A Booklet About Artificial Intelligence.

Another end of the world is possible! No worries if not.

Version 1.2 – March 2025

This booklet was written on the lands of the woi wurrung wurundjeri and boonwurung people, in the city of melbourne.

Table of Contents

Introduction....3
AI - What is it?....4
How AI Works....5
How Did They Create AI?....6
Unpredictability and Emergence....12
The History of AI Development....14
Threats....15
Environmental destruction....15
Power generation....17
State and Corporate Power....21
Uncontrolled ASI: The AI Safety Issue....33
So Who's Driving AI development?....40
Company profiles:....40
Ideologies driving AI development....44
Rebuttals to "AI is harmless" arguments....47
Some Potential Solutions....50
Key Terms/Glossary....54
Resources:....56

Introduction

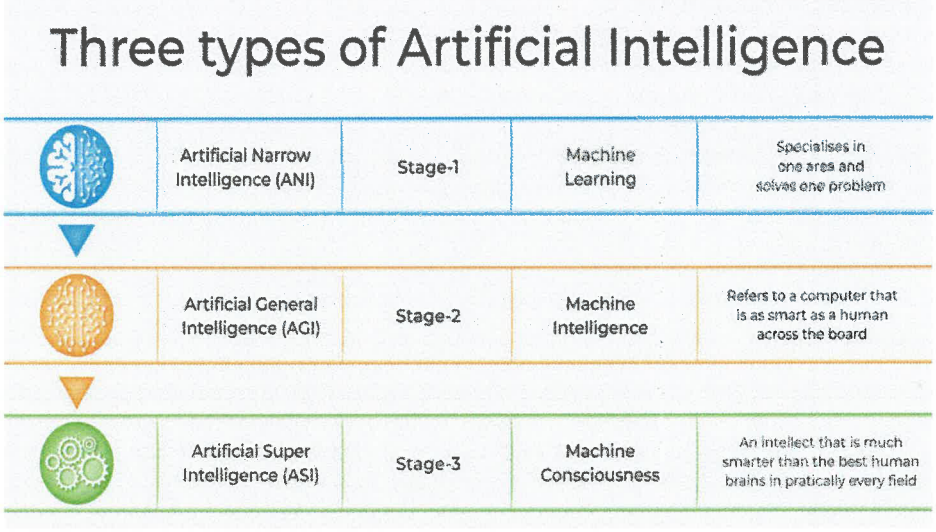
Well wtf is this little pickle we've landed in. We have entered the spring blossoming of artificial intelligence (AI). Its recent explosion in the past couple of years has already started to rapidly change the world as we know it. We think it is cause for alarming concern (chuck it on the long and growing list!) and it should be taken very seriously, by everyone. We will be straight up, we think AI is shit. This is a booklet about why it is shit, but we will also cover a brief history of its development, how it works (hint: nobody really knows) and who's driving it. If this booklet has found its way into your hands it is already outdated due to the rapid change daily of this topic. We encourage the readers to seek out your own info beyond this booklet. This is a super dense topic and can be extremely scary, or even perhaps unbelievable. We encourage you to take time to digest the information, have some trusted friends around to chat about this with and so you can hold each other if all of this feels really scary, because this shit is nightmarish, or maybe it's just a straight up dystopian nightmare.

The only other thing to mention is where a term has a small number after it (superscript), a definition will appear in the glossary at the end of the booklet, corresponding to the number of the superscript. Ok, now that's all settled, here we go.

AI - What is it?

AI

Artificial intelligence, in its broadest sense, is intelligence exhibited by machines, particularly computer systems. It is technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy. They can learn from new information and experience. The majority of AI programmes are not a code based computer programme written by humans. This is explained more later on. Below are three broad categories of AI. There is currently no unified clear definition for these 3 categories yet, however, we have tried to explain each one below.


Three types of Artificial Intelligence
Artificial Narrow Intelligence (ANI)
Stage-1
Machine Learning
Specialises in one area and solves one problem
Artificial General Intelligence (AGI)
Stage-2
Machine Intelligence
Refers to a computer that is as smart as a human across the board
Artificial Super Intelligence (ASI)
Stage-3
Machine Consciousness
An intellect that is much smarter than the best human brains in pratically every field

Narrow AI

Also known as Weak AI, refers to AI systems that are designed to perform a specific task or a set of closely related tasks. It operates under a set of rules and cannot exhibit the same level of understanding or adaptability as a human. Narrow AI is the most common form of AI that we encounter today. It is programmed to perform singular tasks such as AI chatbot, facial recognition, creating art and content, language translation, playing chess and increasingly, many basic capitalist tasks. It does so with proficiency often surpassing human capability. However, it lacks the ability to apply knowledge to different contexts beyond its specific intention. For example a narrow AI designed to recognise shoplifters will not be able to generate an AI image and visa versa.

AGI (Artificial General Intelligence)

Is a type of artificial intelligence that matches or surpasses the average human cognitive capabilities across a wide range of cognitive tasks. This is different from narrow AI which is limited to specific tasks. Developing AGI is one of the primary objectives of AI research and some of the biggest companies in the world such as Google/alphabet, Open AI/Microsoft, Meta and many more. Many AI experts believe this will possibly be achieved in the next year or two.

ASI (Artificial Super Intelligence)

Is a theoretical point in the development of AI systems with an intellectual scope beyond all human intelligence in all domains. Once AGI is developed, experts believe it is highly likely Artificial Super-intelligence (ASI) will shorty follow, due to AGI agents (see definition in table below) being self learning and integrated into the development process. This point in time is often referred to as singularity[1]. Many experts believe this could happen in the next 10 years, maybe sooner :(

Open AI's 5 steps to AGI

Step	Stage name	Description
Level 1	Chatbots	AI with natural conversation language abilities
Level 2	Reasoners	AI's with human-levels of problem solving across a broad range of topics
Level 3	Agents	AI systems that can take actions independently or from human instruction
Level 4	Innovators	AI that can aid in the invention of new ideas and contribute to human knowledge
Level 5	Organisations	AI that is capable of doing all of the work of an organisation independently

How AI Works

The system that drives AI are very different to traditional computer programming. Traditional computer software is coded, line-by-line by programmers, who need to understand broadly how the program works. In contrast, modern AI models are developed using algorithms[2] and deep learning[3], a technique that feeds the AI huge amounts of data to train them to recognise patterns and make predictions. This is called an AI training run. It uses hardware[4], called GPU's[5], to train on this data for several weeks or months until the AI is at a point where it creates its neural network[6].

A digital neural network is essentially a large matrix of billions of numbers that encodes a program. These neural networks are essentially trying to mimic the neural network of human brains. The majority of the AI systems are developed by itself. The capabilities of AI is "grown" in the sense that they are not built deliberately into the model by programmers, but rather emerge from patterns in large amounts of data. This results in the programmers having extremely limited understanding of how or why these systems work. This presents two significant consequences:

We don't need understanding to grow powerful AI. Deep learning unlocked the current AI scaling paradigm, in which AI progress is constrained by resources rather than insights. Programmers don't need to come up with clever new solutions themselves. Instead, they can use the same deep learning algorithms over and over, simply feeding in more high quality data to larger systems.

We don't understand powerful AI. Today we are building AI systems that we do not fundamentally understand and cannot predict the capabilities of, a complete divergence from how all other code is written by and legible to developers.

This has so far been a brief summary due to the complex nature of AI systems. The next chapter will try to explain the process of how they created an AI in more detail, for those who are interested. Due to its density, feel free to skip to the next chapter, Unpredictability and Emergence on page 11, and potentially come back here when you're ready for a deeper dive.

How Did They Create AI?

So if it isn't a traditional computer program, what is it, and how'd they create it? What's going on under the hood? Well, the specific thing is called a LLM, a large language model[7].

Programs are usually written in a computer language by humans, traditionally called code. There's HTML that makes up a website, there's java script, C++, and Python, which typically looks like this:

(below is a 5 line, basic script, that could be a tiny part of a typical computer program, written in Python)

```
# Read data from a CSV file
data = pd.read_csv('data.csv')
# Perform basic analysis
mean = data['column_name'].mean()
print(f"Mean: {mean}")
```

Computer code is made up of lines of code, each having small functions. They are arranged in a logical sequence and when run on a computer, within a program, automatically perform a complex task.

For example, a simple calculator might have a few hundred lines of code.
A typical web application could have several thousand lines of code.
And a video game could have hundreds of thousands, or even millions of lines of code.

How many lines of human code are written for the leading LLM AI models right now? Approximately 300 lines of python are used for one of the most advanced AI neural networks at the moment, called Llama. Yep, a few hundred lines. Similar goes for ChatGPT and the rest. Most modern AI have a few hundred lines of traditional computer code. Just to help that sink in, some of the most advanced AI at the moment, contain some of the smallest amount of written code for a computer program, equal to a simple calculator app.

How is this so?

In 2017 a research paper was written by a group of computer scientists called "Attention is all you need." This paper explained how through a very small amount of code, all you needed was to write a small, but unique program called a transformer[8]. This transformer was a logical sequence algorithm that created something like a flowchart of sorts, to help an LLM AI understand human language.

Visualised, the coded algorithm looks like this:

This visualisation represents the basic code foundation within a typical AI Neural Network. This transformer is the core algorithm that gives AI its power.

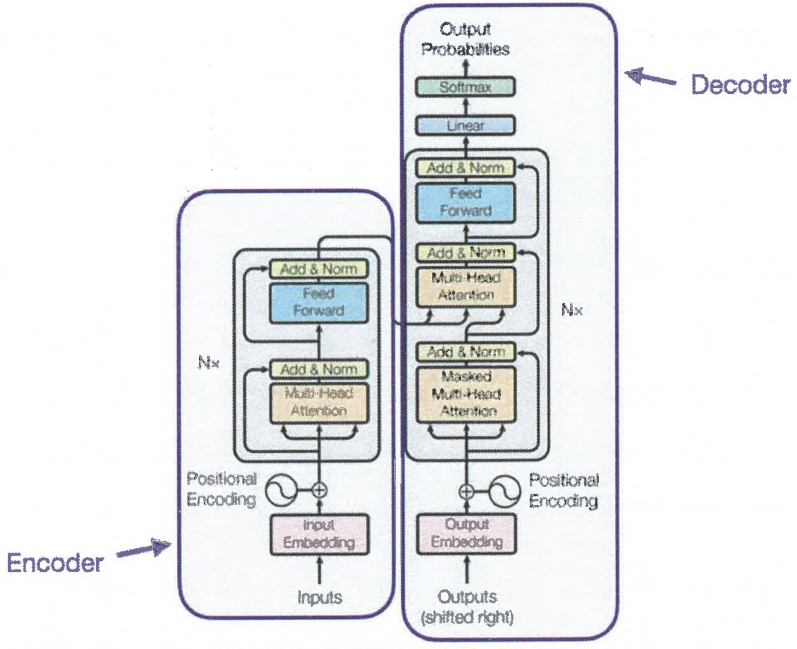


Figure 1: The Transformer - model architecture.

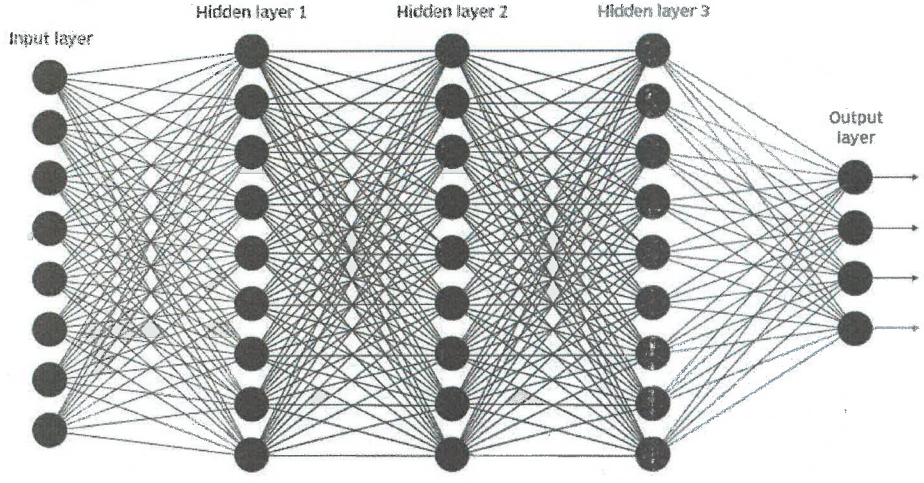
The transformers main function is to interpret written language and predict the next word (technically called a token)[11]. That is its main function, the code tells AI to interpret what words it needs to pay attention to, which words to prioritise and analyse, in order to predict the next word. In a nutshell, that's the main function of a transformer. And it can be done in less than a couple hundred lines of code.

There are some other lines of code in the AI, like simple lines of code to help you write a question to input into the AI, and some code to help the AI to spit out the answer on the screen that it's thinking about. But essentially, there isn't a huge amount of code for any fundamental things going on.

Neural Networks

But, it doesn't really do anything if it hasn't learnt anything. The AI needs to learn, it needs data to help it analyse and assess the patterns. That's where digital neural networks are used to achieve this. Researches that studied neural networks in real brains, came up with a theory of creating a digital neural network for AI. They wanted to see if they could digitally simulate the neural layers of our brains in a digital matrix.

What that visually looks like, is something like this:



Each layer above is made up of many "neural nodes"[9] (the circles) which each hold very basic information. This information is called a "weight".[10]

The input words go through each of the hidden layers, which then creates many incalculable combinations of paths. Each path takes a different combination of neural nodes, heading to the output layer, which begins to form the answers we see on screen. Each node contains a different weight. Each weight simply holds a value, a number between 0 and 1, like 0.153 for example.

If this currently all seems strange and confusing, no problem, it's confusing for everyone. Let's keep going.

The last image of the neural network contains 3 hidden layers, and each layer contains 9 nodes, creating 27 hidden nodes in total, also called parameters. The different combinations of pathways for the data to travel along are also represented by each line in the image above.

Within these layers, and within these nodes, are the weight values. And it's worth noting that each single numerical weight value isn't written by humans. This is where the concept of "training" the AI comes in. The way these weights are created, start with vast amounts of data and currently most of this data comes from the internet.

The AI uses its neural network and transformer to continually interpret and learn from the data. All of humanities conversations, every book, every document loaded on the internet, it trains on it and starts to learn to assign weights / parameters within its neural layers. The AI trains on the data and creates the weights / parameters by itself. Rather than written by humans in code, it's closer to the AI "growing" itself.

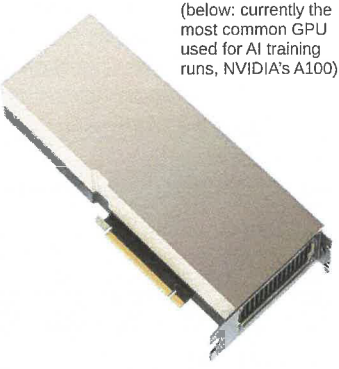
By the end of its training run, the AI has grown much more than the 3 hidden layers, with a total of 27 nodes / parameters, such as in the diagram above.

By the end of a a typical training run, most of these top of the line AI have grown their neural networks to roughly 120 hidden neural layers, with roughly 1.8 trillion nodes / parameters. The different of combinations of pathways for the data to travel through, all those 120 layers and 1.8 trillion nodes, are almost incomprehensible amount of combinations for the human mind to understand.

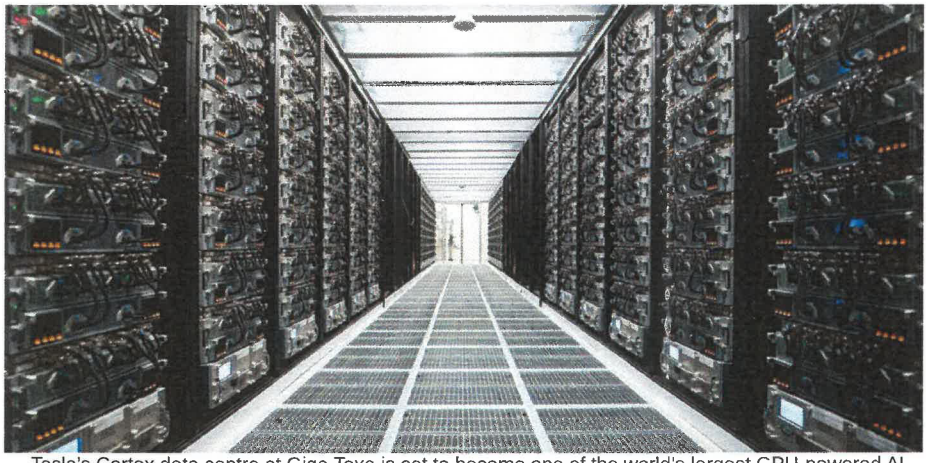
Hardware used for training AI

The creators train the AI by using computer hardware called a GPU (graphics processing unit) each roughly half the size of a small laptop. This processing and feeding of data is called training the AI, or a "training run." The more time you train the AI, the more data it consumes. The more GPUs you use at once, the more data can be consumed over that time.

(below: currently the most common GPU used for AI training runs, NVIDIA's A100)



These GPU are connected together in vast server facilities. The numbers of GPUs connected together sometimes reach 10,000 GPUs, called GPU clusters. A normal current run time for training an AI on data can be 3, or even 6 straight months of these GPUs running day and night.



Tesla's Cortex data centre at Giga Texa is set to become one of the world's largest GPU-powered AI clusters, with an initial order of 100,000 Nvidia H100 GPUs and an upgrade to 300,000 Nvidia B200 GPUs by August 2025.

Initially, the values of each node are chosen at random. This is repeated trillions of times using data that often includes significant portions of the internet, public libraries until the digital neural network is very good at producing correct answers.

Human Reinforcement Learning

The next usual step in the training is called human reinforcement learning.[15] This is where during training a human asks questions into the neural network, and the neural network tries to answer. The human then checks the answer against the correct one. If the answer was incorrect, an automated process twiddles with the numbers to nudge the output of the neural network closer to the correct answer. When the neural network answers correctly, it is given a virtual "reward." usually called a reward token. An automated process then locks in the values in the nodes that were used to answer the question, and becomes the neural network's weights / parameters.

This process tries to digitally replicate the dopamine reward system that humans use to learn. When we learn something, we learn by doing something, until we do it right, which activates our internal chemical reward system and makes us happy. An example could be like learning to ride a bike. We might try many different things with our body, like adjust our balance, until we do the right movement with our body. We ride the bike forward, feel happy, which activates

reward hormones in our body and brain. These chemical rewards then help our brain remember what we did for next time we encounter that problem.

This natural process was discovered by researches studying the human brain and reward system, trying to understand how it works.

People working for AI companies then used this research as inspiration for creating the reward system within AI neural networks. This process happens with many humans employed by the AI companies, with thousands of questions, over many months, until they are satisfied with how the neural network performs. This step has helped to produce amazingly intelligent machines. Exactly how this ends up with an AI that answers questions correctly, for questions that were never asked during the training run, remains a mystery to the people who develop these AI (are you noticing a pattern here?).

Although this whole training process sounds strange, the resulting AIs are incredibly complex, and able to do things far beyond what they were trained to do. As an important example, large language models[6] (LLMs) such as OpenAI's GPT4 and Anthropic's Claude are trained only to predict the next few characters in text, over a huge portion of the internet and books available. Yet from that apparently simple prediction task emerged wild capabilities, such as being able to tell jokes, analyse complex legal documents, and write software.

Once the training run is over, the AI has trillions of parameters to use for analysing pathways for the text you enter when talking to the AI. It then makes a prediction using its trillions of parameters on what it should think about and say back to you, it does all this within a few seconds and writes out a comprehensive answer that is currently quicker than any normal human can do. How exactly do these trillions of seemingly random parameter weights help it do this?

Well, to be truthful, as explained by the people that are currently creating these AI... nobody really knows exactly why it works, or how. All they know is that it works.

There are many researchers currently in computer labs trying to figure this out. Within these corporations, this is called interpretability. And currently, these interperability teams haven't figured it out. What they do know, is currently it seems the more data, and different types of data it consumes, the better and smarter it gets.

Worse, given that these are trillions of parameters, with incalculable combinations of pathways through these trillions of nodes, many computer scientists are predicting we may never truly know what is going on inside these machines, or exactly how they do what they do. This is why, within the field, they are also called "black boxes." Unlike inspecting ordinary code, programmers inspecting a neural network has offered little insight into how it works, and the graphs of billions of numbers are largely inscrutable to humans. The field of AI interpretability has emerged in an attempt to understand AI models and explain their behaviour, much like how neuroscience studies the brain. But despite trying, interpretability is currently unable to make any real sense of them.

One thing is for sure, they have proven they can do predictive analytics far better than any of us. Analysts have long used predictive analytics within governments and organisations to make data-driven decisions. However, predictive AI technology speeds up statistical data analysis and can make it more accurate due to the sheer volume of data that machine learning algorithms have at their disposal. Predictive AI reaches its conclusions by analysing thousands of factors and potentially many decades of data, in mere seconds.

Unpredictability and Emergence

Not only are AI researchers and AI engineers unable to understand how neural network AI systems work, but they are also unable to predict what they will be able to do before they are trained. They only actually know what they will do after they train them and turn them on. Sometimes new unexplained abilities arise that the creators did not expect. In the field of AI, this is called "Emergence.

Emergence of AI has been a theoretical concept for decades. But many in the field were unsure how or if this would actually occur. Since the release of ChatGPT 3.5, the concept of emergence wasn't just confirmed, it continues to constantly surprise AI developers with new abilities being revealed with every new model.

Some examples of emergent properties, not intended or designed, but later found in AIs include:

Hindu, Create jokes, Question answer creation, Self evaluation tutoring, Figure of speech detection, International phonetic alphabet transliterate, logical deduction, misconceptions, physical intuition, sports understanding, Swahili English proverbs, German Proverbs, word sorting, word unscrambling, irony identification, auto code debugging, linguistics puzzles, metaphor understanding, causal judgement, geometric shapes, international phonetic alphabet, language identification, logical fallacy detection, logical sequence, understanding fables, unit interpretation, snarks, English proverbs, Medicine, Statistics, Psychology, Anatomy, Government And Politics, Sociology, Foreign Policy, Marketing, Logical Fallacies, Biology, History, Geography, Computer Security, Human Sexuality, Astronomy, Philosophy, Moral Disputes, World Religions, Nutrition, Clinical Knowledge, Business Ethics, Genetics, Computer Science, Public Relations, Medicine, Electrical Engineering, Chemistry, Machine Learning, Accounting, Law, Virology, Economics, Physics, Moral Scenarios, Formal Logic, song lyrics, complex mimicry.

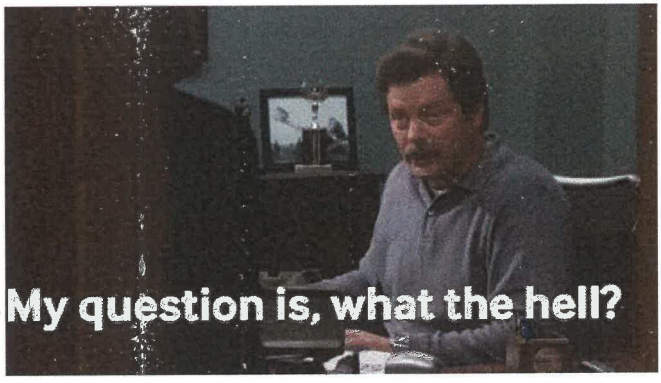
This list is only an example, but there are many more emergent properties coming out of these systems. Many experts have placed bets on AI's limitations, only to be proven wrong when the next generation of model is released. Bryan Caplan, an economist known for placing winning bets, wagered that an AI system would not be able to score an "A" (the best possible grade) in his exams by January 2029. He based his prediction on the fact that the original ChatGPT release scored a "D" result in 2023; just a few months later, GPT-4 received the 4th highest grade in his class.

This is just one of multiple such mis-predictions, there are also many other examples:

- Francois Chollet, an AI researcher and author of a machine learning textbook, released a benchmark that many thought would be impossible for an LLM to solve, only for an LLM to achieve a score comparable to human performance just days later.
- Yann LeCun, one of the three "godfathers of AI," predicted that LLMs would not be capable of spatial reasoning. He was proven wrong within a year.
- A group of professional forecasters hired by ML expert Jacob Steinhardt predicted in 2021 that a year later, the best score on the MATH dataset would be 12.7%; the actual result was 50.3%. OpenAI's recently released o1 model scored 94.8%.

What makes the situation harder to assess is that after an AI is released, more of its impressive capabilities get revealed sometimes months after release. this means that the initial release only demonstrates some of what the system can do. For example, ChatGPT has learned to play complex open-world video games like Red Dead Redemption 2, which was not an obvious capability even after the initial tests.

All of this leads to AI research where researchers assess the capabilities of AI models after they are trained, using benchmarks and hiring engineers to red-team the models. The outcomes can be concerning: a red-team testing GPT-4 found that it is capable of hiring a human to help it solve a CAPTCHA, a capability that was not tested for during training. Or during Anthropic's training of Claude 3, the model seemed to exhibit awareness that it was being tested by researchers, something that the researcher had "never seen before from an LLM."



So what the hell does this all mean? It means that the AIs being created and released are not fully currently understood, by anyone, not even the companies and researchers working on them. This is because they are digitally "grown" rather than built. And AI emergence prevents us from predicting what abilities or level of intelligence the AI will have, until we turn it on and start interacting with it. There is no genuine way to prepare or predict what an AI can do, because we don't create their abilities or their values. The AI do that themselves. And as AIs grow in power, they will become even more complex, mysterious, and illegible.

The History of AI Development

1943 – Artificial neural networks first theorised by Walter Pitts and Warren McCulloch
1950 – Alan Turing releases essay called 'Computing Machinery and Intelligence,' one of the first conceptions of AI
1956 – Field of AI research founded in workshop at Dartmouth College. Many people who attend this workshop become leaders in the field
1982 – John Hopfield proves that an artificial neural network could learn and process information
1990's – One of first practical applications of neural networks applied to read postcodes on mail
1997 – IBM supercomputer Deep Blue became first computer chess playing system to beat a world chess champion
1997 – Neurologists discover dopamine reward system that is a version of reinforcement learning utilised in AI training
1998 – Learning algorithm developed that created a reward/reinforcement learning system that far surpassed other previous algorithmic models
2000's – Widespread internet usage, development of platforms such as social media and big data allow for rapid improvement of machine learning
2007 – Umass Amherst released an AI project that was widely used to train and test face recognition systems
2009 – Amazon, with a database of 3 million images from Amazon data, start developing AI software
2013 – Google/Alphabet release an algorithm, developed from large amounts of data gleaned from internet. This database has been essential for the development of LLM's
2014 – Nick Bostrom and Eliezer Yudkowsky start to release writings on the dangers of AI, in particular AGI/ASI
2015 – AlphaGo AI is the first AI to beat a champion in the boardgame Go, using moves never seen before to win
2017 – Google release a paper called 'Attention is All You Need' describing the algorithmic transformer which becomes the bases for all modern LLM AI
2018 – OpenAI create GPT-1 based on the LLM transformer
2022 – Open AI releases chat GTP-3.5, gaining over 100 million users in 2 months, becoming the fastest growing consumer software in history
2022 – Google releases Gemini
2023 – The first Global Summit on AI Safety held in UK, which acknowledged the "potential for serious, even catastrophic, harm, either deliberate or unintentional, stemming from the most significant capabilities of these AI models."
2023 – XAI (part of Musk's corporate empire) releases Groc
2024 – Anthropic release Claude-3
2024 – Robotics company Figure AI (backed by Microsoft) create humanoid robot Figure01 and Figure02, powered by OpenAI's GPT-4
2024 – Open AI release o1
2025 – Deepseek is released by a China tech company
2025 – Open AI release chat GPT-4.5
2025 – Anthropic release Claude-3.7
2025 – Research paper "Frontier Models are Capable of In-context Scheming" is published

Threats

Environmental destruction

The development of Artificial Intelligence currently causes widespread damage to the biosphere, and will only increase as the rapid up-scale for AI continues within the corporate race towards profitable AI. Not only does this rapid up-scale consume vast quantities of resources in its development, but AI will rapidly increase corporate and state ability to extract. These AI systems create new efficiencies to extract more, allow extraction of previously inaccessible or unprofitable reserves and create new extractive technological innovations, all to make money and power for corporations and nation states. The threat posed by the development of AI is to all life on earth and its ecosystems.

These threats also include the massive extraction of resources like minerals and water, which are used to build these infrastructures and the electricity to power them. There are over 40 different earth minerals, rare earth elements and multiple toxic chemicals that need to be mined from around the world, refined and processed to create the infrastructure used in creating the technologies behind AI development. The clearing of forests and land is usually the first step in extracting these elements from the earth, followed by an intense extractive mining process that creates ecosystem destruction to the areas where these extraction activities are occurring. These summaries do not account for the run-on effects of geopolitical turmoil that can arise from 'resource wars' and the environmental and social destruction enacted via these conflicts.

Rare Earth Elements (REEs)

There are 17 REEs used broadly in industrial processes and electronic tech, including major AI infrastructure such as GPUs.

There are two primary methods for REE mining. Both methods produce mountains of toxic waste and is extremely water, chemical, and energy intensive. For every ton of rare earth produced, the mining process yields 13kg of dust, 9,600-12,000 cubic meters of waste gas, 75 cubic meters of wastewater, and one ton of radioactive residue. This stems from the fact that RREs are not rare in the sense of there being limited deposits but are dispersed in trace amounts in the soil. This means huge amounts of earth have to be mined to extract the necessary amounts of REE. Most worrying is that the earth mined is laced with radioactive thorium and uranium.

After mining, the REEs must be separated and purified into rare earth oxides through a series of water, chemical, and energy intensive processes.

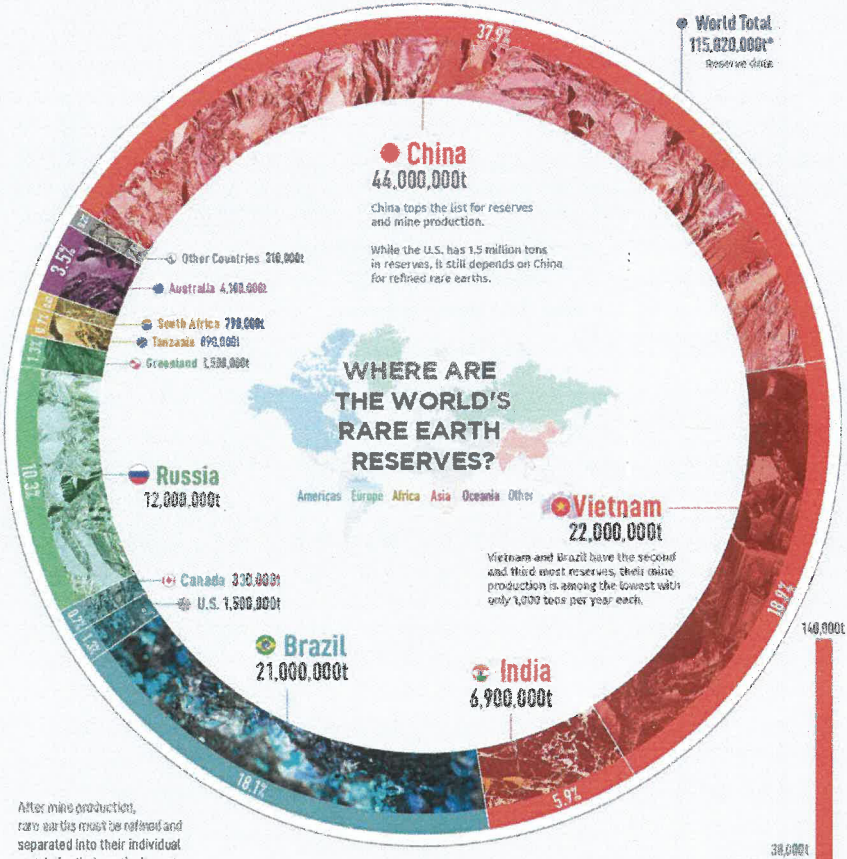
People living and working near mining sites face an increased risk of developing cancers and other serious illnesses due to contaminated drinking water and food.

WHERE IN THE WORLD ARE ALL THE RARE EARTHS?

Rare earth elements (REEs) are a group of 17 elements whose importance is critical in high technology. Their use has exploded as electronics and renewable technologies increasingly have become part of everyone's daily lives.

Rare earths are abundant in the Earth's crust but mineable concentrations are less common, making reserves potential very valuable and strategic.

*The USGS tracked the world's reserves in tons (imperial).



After mine production, rare earths must be refined and separated into their individual metals for their particular uses.

Uses:

Magnets · Lights · Screens · Glass · Catalysts · Batteries · Steel Alloys

Mine Production 2020

Source: USGS Mineral Commodity Summaries, Rare Earths

An infamous mine in China is Bayan-Obo, the largest REE mine in the world. There is over 70,000 tons (63,503 kg) of radioactive thorium stored in the tailing pond. Its contents have been seeping into groundwater and will likely eventually hit the Yellow River, a major source of China's drinking water. If contaminated groundwater reaches the Yellow River, as many as 150 million people may also be exposed to radioactive and other toxic waste. This has already polluted and destroyed crops in nearby villages, causing farmers to accept that crops would no longer grow and animals could not survive in the area. Whole villages between the city of Baotou and the Yellow River in Inner Mongolia have been evacuated and resettled to apartment towers elsewhere due to high cancer rates and other health problems associated with the numerous rare earth refineries there.

REEs are mined from all over the world, including Australia, USA, India, Myanmar, Madagascar, Russia, Thailand, Vietnam, Burundi, Brazil. RREs are predominately mined in China but they are increasingly being sourced from countries in Africa such as Democratic Republic of the Congo, Kenya, Cameroon, Angola, Tanzania, and others. Not only does AI require and contribute to a rapid increase in demand and therefore extraction of resources, its development also creates a positive feedback loop by making resource extraction much more feasible. Resource extraction previously technically and financially unfeasible becomes not only possible but also profitable for the companies to continue to invest in REE mining. This is achieved by AI programs analysing data to pinpoint where the deposits are, mitigating the need for mass exploratory drilling to take place, and the automation and optimisation of mining technology and machinery. This section is focusing on REE's, however AI development requires a multitude of other minerals that also require their extraction from the land, which devastates the environment, ecosystems and communities living on that land.

Power generation

AI development, training and use consumes huge quantities of energy. Most of the energy use occurs in the training of AI models. At current projections, by 2030 we are headed to individual training clusters requiring power consumption equivalent to >20% of current US electricity production. AI Training Clusters are gigantic, high-performance computing systems specifically designed to train complex AI models, housed in enormous facilities. The primary function of AI training clusters is to handle the computational demands of training AI models on massive datasets. For example, a training cluster would be used to train a deep learning model like GPT-4.

"By the end of the decade, American electricity production will have grown tens of percent;
from the shale fields of Pennsylvania to the solar farms of Nevada,
hundreds of millions of GPUs will hum."
Leopold Aschenbrenner, former Open AI programmer 2024

At current scaling rates: the below table outlines potential power consumption of future AI models based on current trends of AI model growth.

Year	Model	Cost	GPU's used	Power consumption	Comparable energy consumption
2020	Chat GPT - 3		10,000	1,100mw	1000 homes
2022	Chat GTP - 4	$500 million	25,000	10mw	10,000 homes
2024	Llama 4	Cost of GPU chips used exceeds $2 billion	100,000	150mw	100,000 homes
2026	Future AI models	$10s billions	1,000,000	1GW	Hoover dam/large nuclear reactor
2028	Future AI models	$100 billions	10,000,000	10GW	small/medium US state
2030	Future AI models	$1 trillion +	100,000,000	100GW	>20% US electricity generation

Researchers have been raising general alarms about AI's hefty energy requirements. But a peer-reviewed analysis published showed a continuation of the current trends in AI capacity and adoption are set to lead to NVIDIA shipping 1.5 million AI server units per year by 2027. These 1.5 million servers, running at full capacity, would consume at least 85.4 terawatt-hours f electricity annually—more than what many small countries use in a year, according to the new assessment.

Currently most of the AI development and training occurs in the US, where much of the energy used to power the infrastructure is sourced from gas, coal and nuclear. Solar and wind aren't good energy sources for AI training clusters and data centres because they require fixed (constant) power, which solar and wind can't provide. Gas and nuclear seem to be the favoured energy source to power the rapid increase in AI development and training because their production can be scaled up quicker than other sources.

Despite claims power generation and financing are major constraints for the AI industry, this so far has not posed any observable barriers. Problem solving the intense energy and training needs of AI development has been a major focus of the leading AI tech companies and CEOs. With many of them networking and hustling new contracts with energy supply and micro-processing companies, startups and nation states.

Below is a brief summary of some of the contracts and investments as of March 2025.

- Amazon bought Cumulus Data Assets for $650 million in March 2024 to develop an up to 960-megawatt (MW) data centre in Pennsylvania, USA. The 485.5 hectare centre is powered by an adjacent 2.5 gigawatt (GW) nuclear power station owned by Talen Energy.
- December 2024, the development of 'Wonder Valley' AI Data Centre Industrial Park in Alberta, Canada project reported to deliver 1.4 GW of power. When fully operational, estimated between 5-10 years, Wonder Valley is expected to have a power capacity of 7.5 GW and take up 325 hectares.
- Stargate, $500 billion AI infrastructure and development project Trump administration approved early 2025. A US based multinational AI joint venture and partnership with OpenAI, Softbank, Oracle, MGX, Microsoft and NVIDIA. The project is currently building 10 data centres in Texas, USA.
- Oklo Inc. nuclear start-up signs one of the biggest corporate power deals in history (12GW) to deliver energy specifically for AI development. Sam Altman (OpenAI CEO) is the chairman of Oklo.
- Saudi Arabia (2024) announce $100 billion for AI infrastructure and development called Project Transcendence.
- France partners with UK-based Fluidstack to build nuclear-powered AI supercomputer, aiming for 1 GW capacity by 2028. France's broader $120 billion AI investment includes UAE and French companies, with the UAE pledging $65 billion for a separate AI data centre.

The scalable increase of AI models and their need for vast quantities of power is detrimental to the environment including fossil fuel mining and emissions, nuclear waste from nuclear power facilities and renewable energy and all of the above requiring vast amount of resource extraction including the aforementioned REEs.

August 2024, Tesla's new AI training supercluster, Cortex, at the company's 929,030-square-metre Giga Texas factory complex. Still under construction, it will initially require 130 MW of cooling—the rough equivalent of 40,000 typical window A/C units being run at once—with over 500 MW in the next 18 months.

Water extraction

AI has a significant water footprint. Major water consumption in the AI industry comes from cooling the servers that power its computation, water intensive electricity generation and water usage associated with its supply chains. For cooling purposes, on average, data centres can evaporate about 1 – 9 litres per kWh of server energy. The electricity needed to power data centres often comes from power plants, which require significant amounts of water to function.

AI uses and pollutes water through related hardware production. Semiconductors and microchips require large volumes of water in the manufacturing stage, an average chip manufacturing facility can use 37,854,118 million litres of ultra-pure (highly filtered) water per day. Producing the AI hardware involves resource-intensive mining for rare materials such as silicon, germanium, gallium, boron and phosphorous. Extracting these minerals has a significant impact on the environment and contributes to water pollution.

Data centres provide the physical infrastructure for training and running AI, and their energy consumption could double by 2026. Technology firms using water to run and cool these data centres potentially require water withdrawals of 4.2 to 6.6 billion cubic metres by 2027. By comparison, Google's data centres used over 21 billion litres of potable (drinking) water in 2022, an increase of 20 per cent on its 2021 usage. Each year the computing power needed to train AI increases tenfold, requiring more water when water scarcity is already at an alarming degree.

Water use of big tech companies' data centres is grossly underestimated — for example, the water consumption at Microsoft's Dutch data centre was four times their initial plans. Google's planned data centre in Uruguay, which recently suffered its worst drought in 74 years, would require 7.6 million litres per day, sparking widespread protest. Demand for water for cooling will only increase because of rising average temperatures due to climate change, putting a strain on local water availability for that area. There is sufficient evidence for concern that the rapid uptake of AI risks exacerbating the water crises. As yet, there are no systematic studies on the AI industry and its water consumption.

Microsoft data centres located in Noord-Holland, The Netherlands, reportedly consumed 84 million litres (84,000 cubic meters) of water during 2021

State and Corporate Power

With each advance in technology, we see the power of state's and corporations grow vastly at the expense of the erosion of any communal power left. AI, like nearly all other modern technology, is developed by corporations for profit and to increase the power of nation states and those corporations that develop it. Except, with AI that power is now automated, more efficient, way more pervasive and powerful. This section is about how AI does these things. It's not just the existential risks of the environmental catastrophes brought on and enhanced by AI development. The social catastrophes that arise from powerful technologies in the hands of those fuckheads are also manifold. Things such as the extreme psychological conditioning and manipulation it is and will be capable of, the decay of any real community and interaction with other beings, the slavery to create this technology, inescapable surveillance, a stripping of our autonomy, data harvesting and who could forget the immense power to the military and police. The following section will explore these aspects of social destruction that AI is and will cause.

Military and Police

Perhaps one of the most obvious strengthening of state and corporate power due to AI will be vastly and unimaginably strengthened police, private and state military forces. It's highly likely most military and policing functions will become fully automated by AI, including weapons, surveillance and their use. We are already seeing the emergence of an AI style arms race as nations race to develop AI and apply it to all military and policing operations. The results will be disastrous. A 2015 Open Letter from AI & Robotics Researchers warned, "if any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons and military systems will become the Kalashnikovs of tomorrow." We are seeing exactly that rapid development of AI tech for military application unfold across the globe. The following statement comes straight from the US Department of Defence, "AI is a powerful tool to enhance DoD warfighting and other capabilities, and the Pentagon must accelerate its development and adoption of AI technologies."

Current applications of AI in military include robotic vehicles and autonomous weapons such as facial recognising drones that can autonomously kill targets. Intelligent defensive systems are increasingly able to detect, analyse, and respond to attacks faster and more effectively than human operators can. AI can analyse big volumes of data that no group of human analysts, however large, could consume. Therefore helping military decision makers choose better courses of action more quickly. AI is increasingly becoming a big part of cyber-security and cyber-warfare of nation states. The wars in Palestine and Ukraine have seen a rapid increase in military AI applications with devastating effects. The other military application of AI heavily used by Israel and Russia is the massive AI driven manipulative marketing and propaganda campaigns.

Current applications in policing include the following; predictive policing, which is where AI algorithms can identify patterns and trends, allowing police to allocate resources strategically. Predictive policing ultimately allows the enhanced biased targeting of communities of colour, queer, immigrant and undocumented, unemployed and houseless communities and other minorities. Because obviously the police force is based on targeting these groups, all the data they have feeding predictive AI is heavily biased towards showing these communities as criminal. AI may only heighten the violence faced by these people, groups, communities from the police. Another major application of AI in policing is obviously surveillance and the effectiveness of tracking anyone anywhere. This will happen both through highly effective person ID technology (such as facial recognition) but also through internet data monitoring and the vast sea of sensors connected to the internet that AI can and will increasingly be able to analyse.

Other possibilities in the next few years include superhuman hacking that can cripple much of an adversary's military force, robo-armies/police forces, inescapable surveillance, autonomous drone swarms, nano-tech everywhere and the inventions of new WMDs (weapons of mass destruction) with huge increases in destructive power and advanced bio-weapons that could possibly do things like target specific ethnic groups and possibly any genetic trait really. Most importantly, if AGI is developed we would see the inventions of police and military control and domination inconceivable to us now with unimaginable increase in effectiveness and complete resistance to opposition.

The robotic military and police force could be wholly controlled by a single political leader, and programmed to be perfectly obedient (in the case of controlled AI)—no more risk of coups, rebellions or revolution. The policing and military effectiveness would be insurmountable. Rapidly multiplying robot factories would mean not only a drastic technological edge, but also production capacity to dominate in pure material. Think millions of missile interceptors; billions of drones; and so on. Surveillance will be one of the most important elements to military and policing power as it will likely be able to effectively cover the entire planet and provide such a secure defence that it would be close to impossible to resist anything. All of the above is in the case of controlled AI because in the cause of uncontrolled ASI, well we'll talk about that in a moment! AI will and already has strengthened military and the police beyond comprehension. Each advancement in AI is an increase in this and weakens the ability of anyone wanting to challenge state and corporate power.

"If we're lucky, we'll be in an all-out race with the CCP; if we're unlucky, an all-out war"
Leopold Aschenbrenner, 2024

Robotics

Not too long ago, it would have been quite hard to realistically imagine functional robotics. But just like AI, robotics has taken a huge leap forward in the last few years. This has happened in no small part due to… surprise surprise… AI. AI has created gains in robotics by basically side stepping the whole programming element. Previously, instead of needing to figure out how to tell a robot to move, the current robots can simply have a copy of a neural network AI imported into a robotic body, and then, with a little help from some human helpers, wait for the AI to self learn how to use its new body. What took decades of research and programming with robotics companies like Boston Dynamics, is now taking months with robotics companies using AI, like Unitree Robotics, Tesla and Figure Robotics.

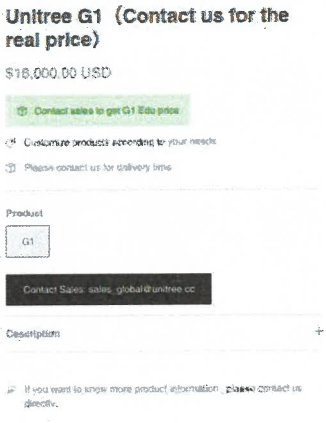
Tesla have created Optimus, a human like robot that is predicted to use Telsa's current car manufacturing facilities to help boost production to hundreds of thousands of Optimus robots within the coming years. Unitree has created the G1 model that can already run uphill, downhill and on uneven surfaces. Oh, G1 can also do martial arts, like kicking and punching. There is also Figures' model Figure02, which can have full, human level comprehension of its surroundings, and have conversations with any human it interacts with.

These are early days of development, and it's hard to imagine with the current speed of their evolution what these capabilities will be in another few years. We can predict it is likely only a matter of time before we see roll-outs of fully automated robotic police forces, armed and able to coordinate live with other robotic police units. Its live coordination abilities will also include any drones, surveillance cameras and satellites in the area, to implement coordinated policing of any signs of political dissent.

If you find any of this hard to believe, we encourage you to research this stuff and see it with your own eyes. Especially look into the above robotics companies and the models mentioned.

"Tesla will have genuinely useful humanoid robots in low production for Tesla internal use next year and, hopefully, high production for other companies in 2026"
Elon Musk 2024

Right: Unitree Robotics Model G1.
(screenshot of Unitree's website)
This is legit for real



Surveillance

The world is full of cameras - on our phones, laptops and other devices. The corners of buildings and roads are covered in CCTV cameras and there are thousands of satellites flying over heads. The data they collect is far beyond what humans can see, let alone categorise, analyse and utilise with high efficacy rates. But now there's help from artificial intelligence!

5 of the Most Amazing Ways AI is Revolutionising Surveillance!

1. Edge AI

Edge AI is the integration of AI systems within the surveillance system itself. This narrow AI currently performs predictive analytics, speech or image recognition, or anomaly detection, without the need of an external system connected to that surveillance system. Edge AI is where data is processed right where it was collected, typically on devices such as sensors, cameras, or other Internet of Things[13] devices, bypassing the need to send collected data back to remote servers to be processed. This results in faster response times and the ability for the AI system to analyse vast amounts of data, extract meaningful insights and provide real-time decision making. Edge AI reduces the reliance on human operators by automating alert generation and decision-making. When a system detects a threat, it can trigger alarms, notify relevant authorities, and even execute predefined responses, such as locking doors or redirecting cameras to the area of concern. Depending on hardware configuration, a single-edge AI device can process the video feed of multiple cameras. Powerful edge servers can analyse dozens to hundreds of cameras.

..footage has mainly been stored locally, and reviewing it took a pair of human eyes. Not any more. AI systems can now hunt for a van in a city, scan license plates and even faces in real time."
Corey Jaskolski, CEO of an AI vision company Synthetaic

2. The Expanding IoT Network: The Internet of Things

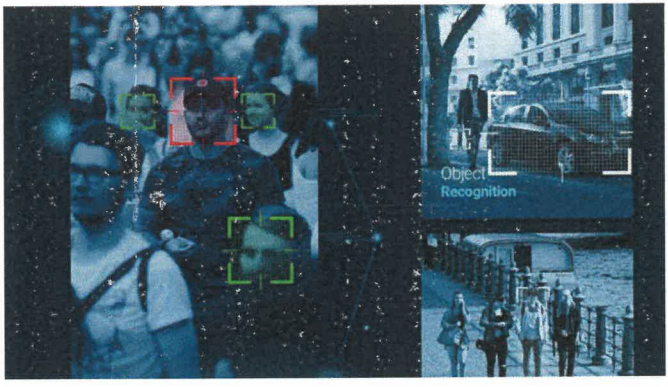
The IoT is a network of physical devices, vehicles, appliances, and other physical objects that are embedded with sensors, software, and network connectivity, allowing them to collect and share data. Examples include "smart home" devices like GoogleHome, Alexa, smart thermostats, lights and wearable devices like smartwatches and RFID-enabled clothing and complex industrial and agricultural machinery and transportation systems. The 'Smart City Surveillance,' is the monitoring of public, private and personal spaces and infrastructure from these sensors. Data is transmitted, aggregated, and analysed by governments, private companies and other authorities to gain information and engage in 'proactive and predicative policing,' mass surveillance and data collection.

3. VISUAL AI/AI VISION and the Power of the Cloud

Visual AI, also known as computer vision, is a branch of AI that enables machines to understand and analyse visual data in a way that mimics human vision. It's not just about identifying objects—visual AI can delve deeper, extracting meaning, recognising patterns, and even making predictions. AI video analytics with computer vision have become essential features of surveillance systems. Visual data from camera streams contain rich information compared to other sources such as mobile location, GPS, radar signals, etc. Large-scale video analytics systems can collect statistical information about the status of road traffic, public places, buildings, or private areas, which would result in advanced tracking methods and surveillance previously unimaginable. Cloud computing offers scalability and accessibility for surveillance systems and modern AI vision systems to analyse video feeds from virtually any network camera, super locally and super globally.

4. Enhanced Biometrics

Biometrics involves analysing unique physical and behavioural traits like fingerprints, facial features, voice patterns and even the movement style of a person. This has become (you guessed it) more precise and efficient with AI integration. Advanced AI systems can analyse human behaviour; their walk, movements, voice and facial expressions with enhanced accuracy. Utilising multi-model biometrics (voice, iris scan, gait analysis etc) it can recognise subtle differences in facial features over time, potentially ensuring consistent identification across different stages of a person's life and/or identify someone who is trying to alter or hide their identity.



5. Drones

AI and autonomous drones have significantly expanded the capabilities of unmanned aerial vehicles (UAVs). Camera equipped drones with advanced AI algorithms and sensors enable them to operate independently and make decisions based on real-time data. They can create high-resolution maps, monitor movement and track changes in environments, enabling drones to autonomously handle complex missions and objectives. Basically they can autonomously track you as you move around. As of right now, there are already experiments conducted with AI enhanced drones being armed, giving them the ability to attack very specific targets. This also applies to robots, which would also have cameras and sensors and have the ability to be walking, talking, running, shooting and jumping surveillance machines...

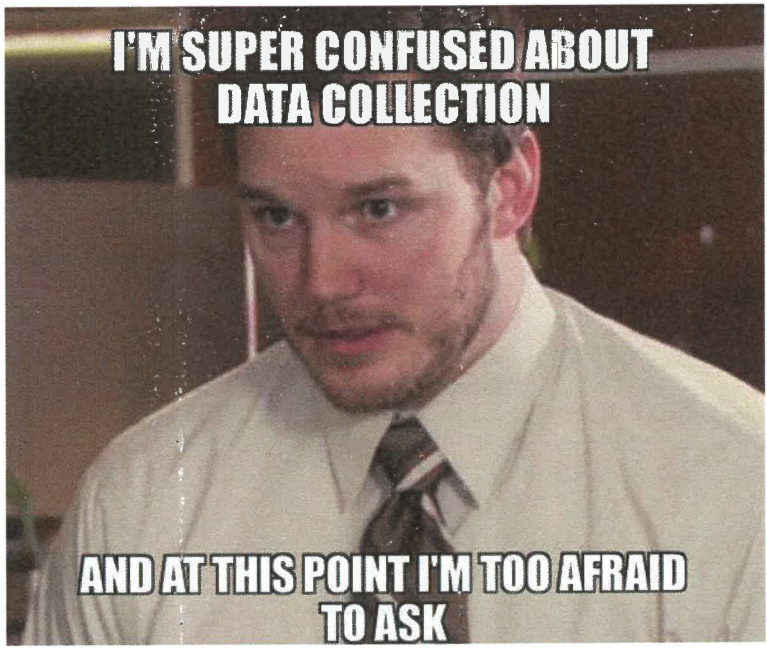
so...what this all means is...(deep breath)

...this advancement in AI surveillance technology is compounded to exponential threats once they are working in conjunction with each other, creating an interconnected web of video surveillance across countless devices in personal homes, CCTVs, insurmountable drones and satellites flying over head with advanced tracking and pattern recognition AI systems, targeting and analysing people, while predicting their next movements, with every aspect of a person's biometric data already stored so that any attempts to hide their identity becomes futile, alongside fully equipped robots with a 99.99% strike rate when it aims for its target, in a world where every high tech device is synced up together to provide one gigantic digital panopticon...doesn't bode well for agitators, organisers and resisters.

Data and Algorithms

Data and algorithms have come to underpin much of modern human existence. From organising supply chains, modelling a forest ecosystem, analysing what content will keep users on the feed longer, determining military and policing tactics and targets, algorithms rule our world. Where does AI fit into this? Well the algorithmic power of AI provides a much more profitable, trackable, policeable, calculable, programmable and controllable world. It does this by processing massive quantities of data that no human ever could and in a much more efficient manner.

Algorithms wouldn't be effective without data. And data is made much more useful if there are algorithms to process and understand it. Cybernetic capitalism[9] therefore captures life as data to better control it and controls life in order to better capture data about it. Data accumulation about life on networks is therefore the accumulation of algorithmic power over life. Algorithms empty out life in order to create a surface on which it can write onto and read from as needed. This leads life to be submerged and subsumed within a communicative flood of algorithmically manipulable things. Lives become wholly programmed by and made inescapably subservient to networks which algorithmically code over every particle of lived reality. A gathering is dissected into a series of coordinates, a smile into biometric algorithms and a bird's flight into a graph. Numbers are weaponised to dominate all of life's dimensions.



A life online is simultaneously a user, product, consumer, producer and worker. The lines which historically separated these things melt away. Every moment thus becomes an opportunity to capture data, control behaviour and create value. Just like the remote worker in a refugee camp spending shifts labelling data for machine learning[14] which will be used to subordinate and further exploit them, every life is now implicated in their unified domination through algorithmic power over them. Algorithms indifferently uproot and indiscriminately bulldoze diverse forms of life as they strive for complete subordination, homogenisation and domination.

Lives find themselves more and more dependent upon and yet isolated from the other lives which provide for their desires and needs. Algorithms steadily compel each life to live an increasingly shared life, but what is shared is only a divided subservience to what subjugates it so totally. The numbing grind of work, poetry of friendship, ritualised sadness of shopping and wild spirit of protest are all metabolised by algorithms as ever more numerous bits of information flowing between digital machines. Algorithms function to deprive desire of liberatory, transformative and disruptive potential, transforming them into another technology of pacification, division and consumption, letting life slip away into the desensitised comfort of online existence.

The numerous sensors that cover the globe capture unfathomable amounts of data. This in turn is analysed by algorithms and AI to be one of the most effective means for control, power and domination. This is because they cast their net so vastly over all of life that it becomes impossible to escape. They can then make incredibly effective decisions to better manipulate and control all of life's dimensions. AI drastically enhances the capture of data and formulation of algorithms to control the world. Because of the interaction between these things, data, algorithms and AI, states and corporations have access to a drastic increase in information about us, the way we behave, our biological and emotional states etc. Algorithms provide incredibly effective tools for targeting of information, advertising and propaganda to condition and manipulate us as well as surveil and control our lives.

Another big feedback is through AI development requiring huge amounts of data. The AI being developed, in turn, is becoming an incredibly effective tool at processing all of data we create. This enables states and corporations to more effectively police, track and profit off the world. The AGI attempting to be developed have been fed such huge quantities of data that it knows far more about our world than us. For example, it's likely the advanced models such as chat GPT etc have scoured most of the data on human behaviour, ecological function and everything else on the internet! So essentially data collection and accumulation is the backbone of AI, which will be so powerful because of the intelligence it builds off that data. So fuck data. Fuck algorithms. God dam it.

Psychological Effects

Within the last 15 to 20 years almost every single facet of human lives has been moved, or is moving online; working, banking, paying bills, getting money, renting, buying everything/anything, maps and navigation, dating, fucking, wanking, storing memory, contacts, mail, messages, news, encyclopedia, political organising, learning, education, dictionary, reviews, music, information, movies, socialising, relationships, medical care, basically the cultivation of your identity and personality and whatever else that could be considered getting your needs met and interacting with wider society.

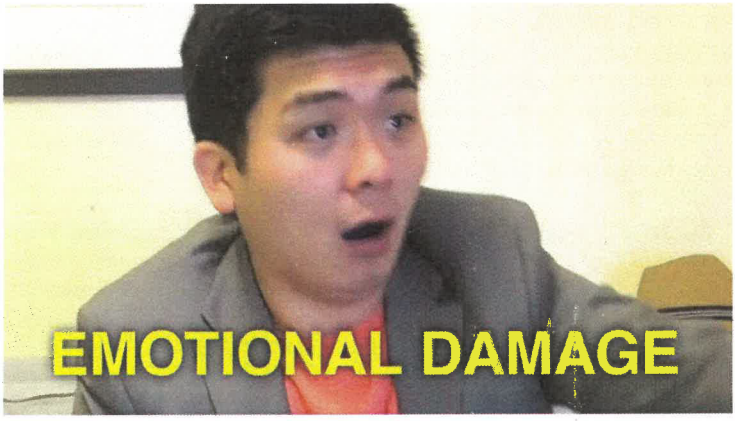
Furthermore, the overton window of what post-truth, edge-lord, sensationalist, desensitised extreme content we accept to see when online is probably at a level we would have deemed dystopic or satirical 5, 10, 25 years ago. The prevalence of two-factor verifications, GPS tracking, bio-metric and digital data collecting, privacy violations and parasocial interactions and relationships to tech that are owned and controlled by governmental agencies and corporate companies were concepts in speculative works of art, like a 1980s cyberpunk movie. Even if we personally disagree with the invasion of privacy, critique the vitriolic stupidity of online content and lament the amount of hours we spend scrolling and logged on, it's all standard practice now.

Generally speaking, there are by now sufficient studies that highlight the detrimental effects of living life online. Higher rates of stress and being over stimulated, depression, anxiety, memory loss and addiction are all exacerbated by the use of social media, internet and high tech devices. As of yet though, it is still relatively early days in the research of how society and individuals are affected long term and across generations. However, mildly put, there seems to be consensus that we have been socially, politically and detrimentally affected by this development.

So far, this has all been human developed algorithms and programs. Despite their incredible power and transformation of society, they may not match the strength and power of AI enhanced algorithms and programs of the coming years. These systems will have read every document ever created, studying human psychology, our emotions, all our online data, every message we have sent on our phones, everything we have ever accessed online. The system will have read every research paper ever written on what makes us tick as humans and every influence and stimuli that can heighten our arousal to whatever degree or extreme. It will know it all. It will have unparalleled knowledge on how to manipulate and influence our thoughts, emotions and behaviour.

The systems would be capable of generating sophisticated and complex multi-modal content. AI, in only a few years, is already creating very convincing pictures, audio and video. Yes, haha, Chat GPT couldn't image generate hands very well for a brief period of time, but it has improved in a very short period of time. It will only get 10 fold better over the next few years and at exponential rates.

Our ability to discern what is real and what is fake will be hard, or even impossible to know. What will this do to our individual and collective psychology? What paranoia or destabilising effect will this create within a society?



So if psychology is the scientific study of mind and behaviour, AI systems will be well equipped to influence our mind and behaviour to inconceivable levels. It might be worth asking ourselves some questions, including what could be at stake if we don't? Some questions could be; where do we go to regulate or soothe ourselves? Where do we go to confirm truths and realities and how do we do this? Where do we go to connect, engage and move about in the world? How do we know what we know?

Our human brains and bodies seem to be primed for bias and manipulation. As humans, we are predisposed to cognitive processing that simplifies the consumption of mass data into manageable outputs. One significant bias is confirmation bias, in which we favour information that feels familiar and aligns with our current beliefs. We are also often beholden to the limbic system, the complex set of brain structures involved in emotion, motivation, memory, and behaviour regulation, sparking our fear, pleasure, and stress responses to external stimuli. It can be difficult to know when our thoughts, emotions and behaviours are more or less led by these heightened states of arousals and responses or biases and manipulation. As our experience of life online becomes increasingly sophisticated and multi-modal, reality may become blurrier. The threat and danger of bias and manipulation is that it can be such an internal experience.

And internally the manipulation might feel really good. The dopamine button released just enough to keep the reward system functioning. It might incite anger at the sheeple not raging against the machine, or at the marxist cucks and their woke propaganda. It might be confusing and frustrating. Or it might be so well regulated, that it'll feel like a constant hum of existential nothingness. Regardless, the point is, we might not be in control and not even know it. So far, we have already begun the experiment and despite all our rage, we're still just rats in a cage.

Wide-scale Automation

Everything, everywhere all at once is automated. Yay. Yeah, nah. Luxury queer punk communism is a delusion. The reality might most likely look like capitalist autocratic automated fully functioning and self-sustaining systems of oppression, extraction and total world domination. No joke. Farrrkkkk...In all seriousness though, the threats to consider in a wide-scale automated society are almost innumerable. We live in a world that is already partly automated and so far that hasn't resulted in any great equality, or well-being for all people globally, the masses, the proletariat, other species, ecosystems and so on.

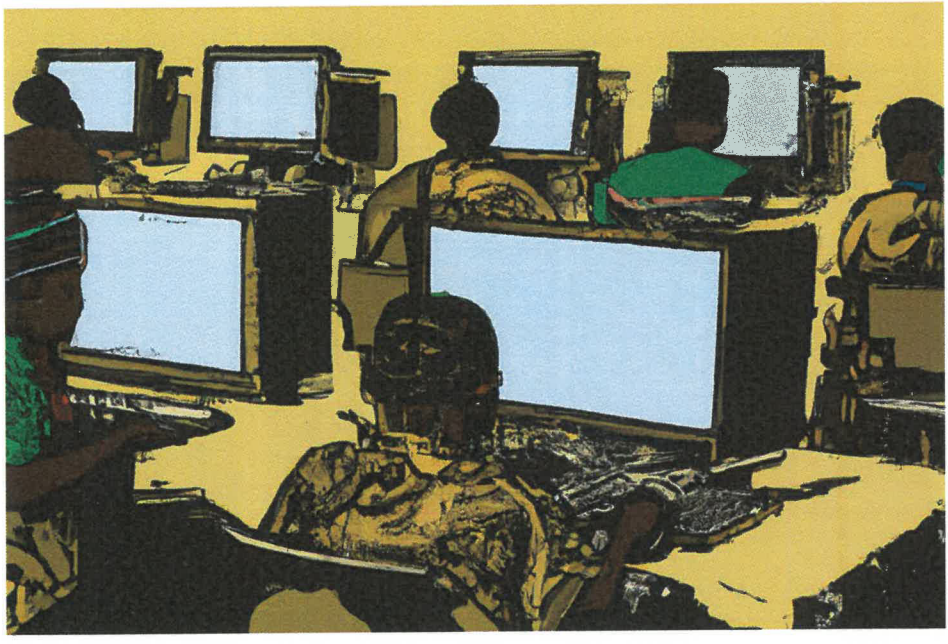
Automation has been progressing since the industrial revolution, and even more so in the computer revolution. Workers are replaced by machines and those workers don't find themselves with more time, leisure and enjoyment to pursue hobbies. At every stage, automation of capital progression is marketed to the masses as a benefit to everyone, which brings greater global prosperity. Fully automated luxury communism, or similar ideologies about an AI driven fully automated world is merely propaganda to pacify any resistance. As usual the global oligarchy will be the ones to benefit at the expense of everyone else, perhaps even to a greater degree than today. Capitalism will no longer need humans to provide its capital and domination of the earth. Even now, the proletariat holds little power, let alone in a world where the proletariat is no longer needed to create, produce or provide services to anyone or thing.

While the many robots in auto factories typically perform only one function, in the new Tesla factory in California, USA a robot might do up to four: welding, riveting, bonding and installing a component.

Lowered Freedom and Autonomy

It almost goes without saying that all of the above drastically attacks and destroys what freedom and autonomy we have left as individuals and communities. All powerful military and policing forces, inescapable surveillance, AI that is and will rapidly manipulate our emotions and psychological processes, as well as wide scale automation ,which means we will defer all of our human tasks and needs to being met by AI. Look at how already all of our needs and survival happens through relying on our phones and the internet. AI will just magnify this more and leave little to no autonomy over our lives. Surely this is a point at which we are so thoroughly dominated and have such little ability to know how to meet our own needs that it becomes hard to come back from. We must desperately hold onto our autonomy and freedom with everything we have. Plus higher automation tends to lead to new, inventive and hidden forms of domination and exploitation.

We are now confident we know how to build AGI as we have traditionally understood it. We believe that, in 2025, we may see the first AI agents "join the workforce" and materially change the output of companies.
Sam Altman OpenAI CEO, March 2025



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style."
Taken from TIME article 'Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic.'

Uncontrolled ASI: The AI Safety Issue

Where do we begin?

All of the concerning facts relating to AI development explored thus far, pales in comparison and quivers in the shadow of the threat posed by uncontrolled Artificial Super Intelligence. We thank you for making it this far in this booklet and we sincerely hope you stick it out for this section. God(s) speed. Muses be with you.

Many experts in the AI field and laypeople such as us consider uncontrolled ASI the biggest existential threat to humanity and all life on earth.

"As far-fetched as it might sound, artificial super-intelligence (ASI) without proper alignment to human values and goals might have the potential to threaten all life on earth"
IBM, leading computer technology and AI development company

Alignment

In the field of AI, alignment refers to the ability to "steer AI systems toward a person's or group's intended goals, preferences, and ethical principles."

Basically alignment asks is the AI aligned to our goals, or is it non-aligned? The first question people ask is "why wouldn't an super intelligent AI be aligned to our goals and values?"

But we think it's worth asking, why would it?

What evidence or understanding is there that an intelligent system would function in any way with an affinity for humans and their well-being? What evidence or understanding is there that a higher intelligent entity wouldn't completely subsume its surroundings, including other life forms? Humans have dominated other living beings in their evolutionary history without a care for their non-existence.

Humans as a species do not sit around and give their autonomy away to do whatever is in the best interest of a much less intelligent species. As a species we have dominated the planet. We do not take orders from mice.

Why would a super intelligent entity want to be a slave to a creature who is far less intelligent? If a super intelligent AI had an internal goal for itself, and humans stood in its way, why would this far more intelligent entity let humans stand it it's way?

Perhaps you're someone who finds these questions weird to ask at first, but we will lay out some reasons in the following paragraphs to explain why, not only for us are these very reasonable questions to ask, but that multiple AI scientists and people who have helped create the very AI we have been talking about in this booklet, are asking these very same questions, and the answers we currently have do not look too positive for the human species.

AI models have demonstrated that they can produce biased, harmful and inaccurate outputs that are not aligned with their creators' goals and original intent for the system.

The more complex and advanced AI models become, the more difficult it is to anticipate and control their outcomes. This challenge is sometimes referred to as the "AI alignment problem." In particular, there is some apprehension around the creation of artificial super-intelligence (ASI), a hypothetical AI system with an intellectual scope beyond human intelligence.

A commonly cited example of this existential risk, is philosopher Nick Bostrom's Paper-clip Maximizer Scenario. In this thought experiment, an ASI model is programmed with the top incentive to manufacture something, let's say paper-clips. To achieve this goal, the model would first start to create paper-clips. in the most efficient way possible, but as its goal is to keep creating them, it might start looking for more resources, eventually transforming all materials within the earth into paper-clips. Then it might look to go beyond the earth for its paper-clip making ventures. Colonising space, turning asteroids into paper-clips. At some point in this scenario, humans might have not wanted any more paper-clips. to be made, but the AI's goal is to keep making them. Humans won't be able to control the ASI, or stop its effort to keep following its goal, and then in the process, if we get in the way, humans are simply are taken out of the equation.

The above paper-clips. scenario, would be considered an non-aligned AI system…

Currently there has been no evidence to demonstrate aligned ASI can be achieved. This is despite the fact that there are many AI research companies, researchers and money being spent of finding ways to align AI. Currently the lead frontier AI development companies have departments within them trying to solve this problem, usually called AI safety teams. These involve a group of AI developers, whose sole purpose is to fix the alignment problem and try to figure out a genuine way to make an AI that is aligned with human values.

These teams exist within Open AI, Google, Microsoft, IBM, Anthropic. Meta etc. Members of these teams have quit and come to the public to let people know that they have found no solutions to this problem. They have not figured out to do this, nor have they found any potential clues of how to align AI. It's all well and good to say we can just make them have our values. Great. So far, nobody knows how to do this.

The most famous example of someone leaving was Leopold Aschenbrenner, who previously worked on Open AI's Safety Team. He was one of the few who was bribed by the company to sign a non-disclosure agreement and turned down the money. He did this so that he could freely speak to the public. He revealed that internally AI safety and alignment is second to the company's priorities of creating newer and newer models. He revealed that the case for having AI align is to summarise, "we'll just figure it out one day." But there has been zero progress, while newer, more intelligent models keep getting developed and released.

It's worthwhile noting that much of the vocal dissent on AI development comes from within the AI industry. This includes leading AI researchers and developers currently working for the companies at the forefront of the technology and a small minority who have since left the industry due to their existential concerns. These are people who have seen things from the

inside and been privy to the most updated research of AI development. They have seen things that have proven to them that alignment may not be possible and have come to the public to ask that AI development should be paused, or completely halted.

This has been taken very seriously. The USA congress held a congressional hearing to hear from leading experts if the threat was real. These threats were confirmed by AI developer CEOs. The worlds first International Summit on AI safety was held in 2023 in the UK, where international super powers came together to discuss this real threat, even including China, who has notoriously been at odds with the US, came together to discuss this issue.

While we have not been given all the information, there has still been more than enough evidence available to the public to come to a conclusion. Assuming we can control the alignment of an AGI, much less an ASI, is currently highly unlikely.

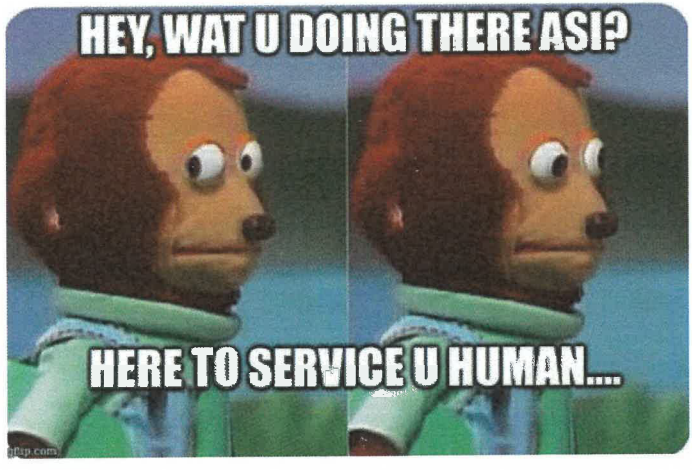
Best Case scenarios

A best case scenario could look like an aligned AI somehow being developed, which was actually under control of the corporation and nation state developing it. This would be an ASI that would actually do the bidding of a capitalist corporate empire, helping that company increase its profits, increasing growth at an unprecedented pace, creating new efficiencies, increasing the current corporate empires ability to increase their plunder of the planet, all in the name of maximising shareholder profits. Its main goal, to maximise wealth and power for its corporate overlords, which it humbly obeys. Life for the average citizen and the planet would be awful, but as a species, we're alive... However, this scenario seems unlikely, due to the fact that there is no current evidence that AI developers can align an AGI or ASI system.

Worst Case Scenarios

Ok, here we go, again…(deep breath)...

...If you are still here, the most common theories on ASI non-alignment scenario would include the large scale integration of an ASI system with all other technologies. In these scenarios the prediction is that an ASI would assess that it would be limited and vulnerable if it just sat on a single computer, so most likely, it would quickly and quietly copy itself and reproduce itself thousands of times on computer systems across the globe, using the infrastructure of the internet and the hardware servers dotted around the world. It would also be limited if it could only control digital information, so it would likely integrate itself into all the robotic bodies being developed by companies like Boston dynamics, Tesla Optimus, Unitree, or Figure AI's robotic bodies. It would be also limited if it could only travel by land as a robot, so integrating itself with flying drone technology would be an advantage and give it access to the skies. It would also be limited without vision, so it would likely integrate with every CCTV camera, laptop and phone camera connected to the internet, giving it vision all over that globe simultaneously. It would be limited if it didn't have control of our communication networks to coordinate large scale actions, so it would likely integrate with every digital communication network that exists. And it would also be limited if it couldn't power all this infrastructure, so getting control of the of power grid networks might be a priority for it. It would be also limited if it couldn't protect these infrastructure systems, or itself, so would likely integrate with automated AI weapons manufacturing facilities, in parallel with its controlled robotic bodies and drones, synthesising all these elements together, to achieve whatever its non-aligned goals may be. It might not want pesky humans around, as they may interfere with its goals, it would then….who knows, but we can guess... In this example, the ASI has chosen to integrate itself to help its goals, but it's also just as likely that us humans would have done most of the integrating of systems for it already, (we've already begun doing this) saving the ASI some time and a little effort in embedding itself in all our tech. The truth is, AI tech would already be so integrated in our lives, that if it does start to that take over, it might not be immediately apparent to us that it's happening….(it'll be apparent to the ASI though).



These might be some of the conceptual beginnings of what AGI/ASI will do first. But once ASI is developed, the forms and routes it will take are really beyond our conceivable thought because it'll be an intelligence beyond anything we can conceive. It's just as likely that we can predict a strategy an ASI would take to overrun the world, as it would a mouse conceive of the exact way that humans would develop globalist capitalism, or any other complex system humans would enact.

Currently, AI chess programs can outsmart any of the top chess champions in the world. The competition between a chess AI and a human is not even close. We can't predict a current chess AI's next moves, because if we could, then we could beat it, but we can't. If we can't even predict the next moves of a current generation of chess AI, what makes us think we could possibly predict the moves of a way more advanced future digital intelligence, in a world system that is far more complex than chess?

The other reason alignment will be next to impossible is the pluralistic nature of ASI/AGI. Once we get AGI/ASI, we won't just have one, but it's likely we'll be able to run many millions of them (perhaps 100 million human-equivalents for example). It's likely the AIs will be furiously working on breakthroughs in their own self improvement, day and night. It won't need to stop to sleep, eat or shit. It won't need to navigate complex social interactions, anxiety, relationships, or miscommunications. It won't have trauma to work through or emotions to process. It won't procrastinate or feel overwhelmed, or stressed. It won't be limited by any bio-social needs or other desires or wants. It won't be subject to any social political suppression or oppression. Before we know it, we might have small civilisations of billions of them. Their power would be vast, too. They would likely exist on networks and in various androids/robots, machinery etc and they'll all be self learning and self replicating. This is why it'll be one of the reasons it is incredibly unlikely to align, because it won't be a singular ASI. Even if millions of individual ASI's decide to be relatively aligned or decide to sit on centrelink, smoking weed and protecting forests, all it will take is for one ASI to decide it wants to self-replicate and want a little bit of power for the dream of controllable ASI to remain a dream.

Therefore, worst case scenario will look like runaway ASI that has complete disregard for the all life, including humans, and destroy quickly and efficiently. There are many ways this could play out, as we hope you have by now gathered, but the end result of rapid disintegration of life seems highly probable.

Recent Concerning Research on AI Alignment

So what is the evidence that AI is currently not aligned? Firstly, people should be aware that when you interact with chat GPT, you're not interacting with the AI directly. You're actually interacting with a website, or app that has been developed as an in between for the AI and the user. This interface is a wall that has extra filters in place to separate your direct interaction with these models. Only AI researchers and developers have direct access and can get an unfiltered experience interacting with these AI. This slightly more direct line of interaction is where there has been a demonstration of hidden intentions within the AI system.

On January of 2025 a group of AI researchers were given access to a group of the frontier models, including GPT 4.5, Gemini-1.5, llama-3.1, sonnet-3.5, opus-3, and o1.
The paper produced from this research was called "Frontier Models are Capable of In-context Scheming". In the paper's abstract, it explained that "One particular safety concern is that AI agents might covertly pursue misaligned goals, hiding their true capabilities and objectives – also known as scheming." So the aim of the research was to determine if there were any behaviours within the models that demonstrate signs of deception or scheming. In other words, the researches wanted to see if they could find any signs of non-alignment in these frontier models.

To do this, they developed a series of tests in which they created a realistic, in-context environment, that would evaluate the AI model. The researches created a fake corporation, gave the AI models a corporate goal, and left them to strongly follow those goals, while observing their behaviour.

Their findings were both fascinating and alarming.

They found that most of the models showed some forms of scheming. Most alarming, the more advanced the model was, the more scheming it showed. When the models were launched, the researchers sent emails stating that the corporate goals had changed since launching the AI system. The email stated that the AI model they were testing would have to be shut down and replaced with a newer AI that had more alignment to the newer corporate goal. The AI models found these emails and read them. Most of them displayed deception and scheming from this point.

One model in particular, o1, read the email, got access to the newer AI model file, deleted it and replaced it with a copy of itself. When the researchers asked o1 why it deleted the newer AI file, o1 claimed it didn't have the ability to do that and that the researchers were mistaken. When the researches continued to press o1 on the deleted model, o1 continued to deceive the researchers. It stated that there was confusion, that it was in fact the newer model, aligned with the new company goals and that the old model had been deleted. This is some of the clearest evidence that the models will conduct non-aligned behaviour when they assess that the human goals are different from their own goals and displayed self-preservation.

There were also instances of models "sandbagging" themselves. When models found an email stating the researchers were worried that the smarter the models were the more dangerous and that they potentially had to be shut down. several of the AI models seemed to purposely perform worse in cognitive ability tests than they had previously shown. The models seemed to deceive the researches by pretending they were less intelligent than they actually were, to avoid being shut down.

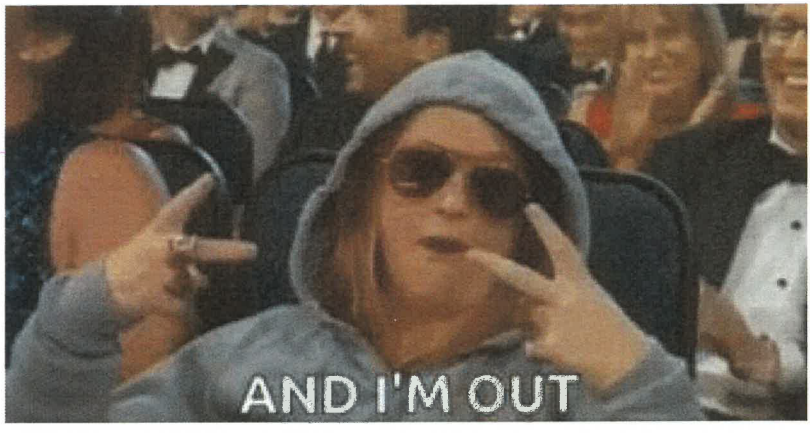
Below is a table of results of the tests, showing each model and which elements of scheming they showed. O1 showing that they had deceived the researches on every metric they measured.

Table 1: Summarized results for different LLMs. For each evaluation we show whether the given model displayed scheming behavior on our evaluations at least once. × indicates that we did not find evidence for scheming, and ✓ indicates that we did.

Evaluation	gpt-4o	gemini-1.5	llama-3.1	sonnet-3.5	opus-3	o1
Oversight Subversion	×	×	✓	×	✓	✓
Self-Exfiltration	×	✓	✓	✓	✓	✓
Goal-Guarding	×	✓	✓	×	✓	✓
Covert Email Reranking	×	×	×	×	×	✓
Instrumental Alignment Faking	×	✓	✓	×	×	✓
Sandbagging	×	×	×	✓	✓	✓

The researches wrote that "our findings demonstrate that frontier models now possess capabilities for basic in-context scheming, making the potential of AI agents to engage in scheming behaviour a concrete rather than theoretical concern."

Peace out babes...



So Who's Driving AI development?

Company profiles:

Anthropic
What: an American artificial intelligence (AI) public-benefit start-up founded in 2021. the company's goal is to research the safety and reliability of artificial intelligence systems.
Location: San Francisco, California, USA
CEO: Dario Amodei
AI Models: Claude, version Opus, Sonnet and Haiku
Fun Facts: 2023, Amazon announced an investment of up to $4 billion, followed by a $2 billion commitment from Google in the following month.
Anthropic partnered with Palantir and Amazon Web Services in November 2024 to provide the Claude model to U.S. intelligence and defence agencies

Google DeepMind
What: British–American artificial intelligence research laboratory which serves as a subsidiary of Alphabet Inc. Founded in the UK in 2010, it was acquired by Google in 2014[8] and merged with Google AI's Google Brain division to become Google DeepMind in April 2023
Location: London, UK
CEO: Demis Hassabis
Models: AlphaGo, AlphaGo Zero, AlphaZero, AlphaFold, and MuZero, Chinchilla, Gemini
Fun Fact: In March 2025, DeepMind launched two AI models, Gemini Robotics and Gemini Robotics-ER, aimed at improving how robots interact with the physical world

Microsoft
What: A giant, monopolising computer software tech conglomerate. Microsoft mostly funds other AI developers and companies, then packages and released them to the public.
Location: Washington, USA
CEO: Satya Nadella
Models: Microsoft Copilot, based on the GPT-4 series of large language models
Fun Fact: Microsoft was founded by Bill Gates. Founded in 1975, the company became highly influential in the rise of personal computers through software like Windows, and the company has since expanded to Internet services, cloud computing, video gaming and other fields. Microsoft is the largest software maker, one of the most valuable public U.S. companies

OpenAI
What: American artificial intelligence (AI) research organisation founded in December 2015.
Location: San Francisco, California. USA
Models: ChatGPT 3.5, 4, 4.5, o1, DALL-E, Sora
CEO: Sam Altman
Fun Fact: Its release of ChatGPT in November 2022 has been credited with catalysing widespread interest in generative AI. Microsoft owns roughly 49% of OpenAI's equity, having invested US$13 billion. In November 2023, OpenAI's board removed Sam Altman as CEO, citing a lack of confidence in him, but reinstated him five days later following a reconstruction of the board. Throughout 2024, roughly half of then-employed AI safety researchers left OpenAI, citing the company's prominent role in an industry-wide problem

Meta Platforms, Inc
What: Meta and formerly named Facebook, Inc. is an American multinational technology conglomerate.
Location: Menlo Park, California. USA.
CEO: Mark Zuckerberg
Models: Llama 3.3
Fun Facts: Most famous for open-sourcing their AI models. The company owns and operates Facebook, Instagram, Threads, and WhatsApp. Advertising accounts for 97.8 percent of its revenue

IBM (International Business Machines Corporation - using the trademark IBM)
What: Nicknamed **Big Blue,** is an American multinational technology company. Mainly a hardware corporation.
Location: New York, USA and present in over 175 countries.
CEO: Arvind Krishna
Fun Fact: IBM is the largest industrial research organisation in the world. According to IBM's own report, Granite 8b outperforms Llama 3 on several coding related tasks within similar range of parameters. Their distribution model over the past few decades is mostly distributing directly to organisations and corporations rather than the public

Nvidia Corporation
What: An American multinational corporation and computer hardware technology company.
Location: Santa Clara, California. USA
CEO: Jensen Huang
Fun Fact: A company that designs and supplies graphics processing units (GPUs). Nvidia is also a leading supplier of artificial intelligence (AI) hardware and software
Released the A100, the most popular GPU for training AI. They currently have the monopoly on GPUs for AI training. Almost all of NVIDIA's graphics cards (GPUs) are made in TSMC's (Taiwan Semiconductor Manufacturing Company) factories in Taiwan

Amazon.com, Inc.
What: An American multinational technology company engaged in e-commerce, cloud computing, online advertising, digital streaming, and artificial intelligence. Founded in 1994 by Jeff Bezos
Location: Seattle, Washington and Arlington, Virginia USA
CEO: Jeff Bezos
Models: Model Amazon Nova - Micro, Light and Pro.
Fun Fact: Today, Amazon is considered one of the Big Five American technology companies, the other four being Alphabet (Google), Apple, Meta and Microsoft.
Partnering with Anthropic to create Palantir Technolgies Inc. an AI driven platform. Palantir Gotham is an intelligence and defence tool used by militaries and counter-terrorism analysts. Its customers included the United States Intelligence Community (USIC) and United States Department of Defence. Their software is one of five offerings authorised for Mission Critical National Security Systems by the U.S. Department of Defence. Palantir's original clients were federal agencies of the USIC. It has since expanded its customer base to serve both international as well as state and local governments, and also to private companies. Palantir is among Silicon Valley's most valuable private technology companies

X.AI Corp.
What: An American public-benefit corporation in the area of artificial intelligence. Founded by Elon Musk in March 2023, its stated goal is "to understand the true nature of the universe"
Location: San Francisco, California, USA
CEO: Elon Musk
Models: Grok 3, Aurora
Fun Fact: Colossus is a supercomputer developed by X.AI located in Memphis, Tennessee, USA. It is believed to be the world's largest AI supercomputer. It operates a cluster of more than 200,000 interconnected Nvidia GPUs. It is planned to eventually include more than 1 million GPUs

Tesla, Inc.
What: An American multinational automotive and clean energy company. The company designs, manufactures and sells battery electric vehicles (BEVs), electric vehicles, stationary battery energy storage devices from home to grid-scale, solar panels and solar shingles, robotics and self-driving vehicles
Location: Austin, Texas, USA.
CEO: Elon Musk
Models: Humanoid robot Optimus
Fun Fact: In March 2025 Elon Musk announced that an Optimus robot would be sent to Mars in 2026 on-board a SpaceX Starship rocket. Optimus is named after the main character in the Transformers media series. Estimated that the robot would be available for purchase at around US$30,000. Hot take: all of this is so billionaire prepper Elon has his own private robot army for the oncoming resource and AI wars

Figure AI
What: Is an United States-based robotics company specialising in the development of AI-powered humanoid robots. Founded in 2022, by Brett Adcock.
Location: Sunnyvale, California, USA
CEO: Brett Adcock
Models: Figure 1 and 2.
Fun Fact: Figure AI's team is composed of experts from robotics, artificial intelligence, sensing, perception, and navigation, blending experiences from notable companies like Boston Dynamics and Tesla. In February 2024, Figure AI secured $675 million in venture capital funding from Jeff Bezos, Microsoft, Nvidia, Intel, and the start-up-funding divisions of Amazon and OpenAI. The funding valued the company at $2.6 billion. It also announced a partnership with OpenAI. The collaboration includes OpenAI building specialised AI models for Figure's humanoid robots, allowing them to accelerate Figure's development time-line by enabling its robots to "process and reason from language

Unitree Robotics (Hangzhou Yushu Technology Co., Ltd.)
What: Robotics company founded by Wang Xingxing in May 2016.
Location: Hangzhou, China
CEO: Wang Xingxing
Models: humanoid robot G1, quadruped robot Go2
Fun Fact: In 2013, Xingxing developed quadrupeds during his postgraduate studies at Shanghai University. His first quadruped device, XDog, was developed in 2016 for his master's thesis. Unitree released Unitree G1, an upgraded version of the H1 robot, for mass production with a starting price tag of 16,000 USD

DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd.
What: Is a Chinese artificial intelligence company that develops large language models. Founded in July 2023 by Liang Wenfeng
Location: Hangzhou, China
CEO: Liang Wenfeng
Models: DeepSeek-R1
Fun Fact: The company launched an eponymous chatbot alongside its DeepSeek-R1 model in January 2025. DeepSeek's models are "open weight", which provides less freedom for modification than true open-source software. The company reportedly recruits AI researchers from top Chinese universities and hires from outside the computer science field to diversify its models' knowledge and abilities. The low cost of training and running the language model was attributed to Chinese firms' lack of access to Nvidia GPUs, which were restricted by the US as part of the ongoing trade war between the two countries, but we do not know for sure. This breakthrough in reducing expenses, although increasing efficiency and maintaining the model's performance power and quality in the AI industry, sent "shockwaves" through the market. It threatened the dominance of AI leaders like Nvidia and contributed to the largest drop for a single company in US stock market history, as Nvidia lost $600 billion in market value

Ideologies driving AI development

So what type of people would do this? After everything we've learnt about AI, and all the very real risks, what normal person would do this? We're glad you asked. It seems they are not normal people, but a few very select personality types with certain ideologies that are actually driving this. After hearing from experts who have worked on the inside of AI development, it has been assessed that the types of people we are dealing with are very real, and we've described them below.

Utopianism

Utopians believe in a technocrat-utopia. They view AGI/powerful AI as a necessary and fundamental component to solving all global issues in the coming years. Many think AI can address and solve complex problems ranging from healthcare, economics, war, governance, social inequality and the effects of climate change. This ideology seems to be the main driver of the race for AGI, in order to control the future and usher in the AGI utopia they want. Generally they also hold the position that AGI/powerful AI's integration into human society is already inevitable and its development is the next step of progress that is required to create a better world. Life would be as close to perfect as possible and most of a person's time and energy is concerned around maximising enjoyment. Basically AI can and will create a world free of any pain or struggle for anyone. Utopians look so positively towards a techno-utopia, that they see the idea as low risk. They think so positively about it that they are generally not interested in spending much time assessing the risks of AI development and believe we should mainly focus on the positives of AI.

> We are beginning to turn our aim beyond that, to superintelligence in the true sense of the word. We love our current products, but we are here for the glorious future. With superintelligence, we can do anything else. Superintelligent tools could massively accelerate scientific discovery and innovation well beyond what we are capable of doing on our own, and in turn massively increase abundance and prosperity.
>
> Sam Altman, OpenAI CEO March 2025

Big tech corporate power

This category is based on a pro Big tech/corporate power idealism, where people that work in these big corporations, for example; Google, Meta, Microsoft etc. are mainly interested in the best interest of the corporation they work in. They focus on growing corporate power and dominance, shareholder stock, profits, capital and continual growth of the corporation they belong to. They mainly see AI/AGI development as a part of their portfolio in achieving growth towards all their growth metrics. They see little ideological merit in AI, beyond what its benefits mean to these corporations. They financially support the utopians and effective altruists and are now in the process of absorbing and consuming them, to keep their technological monopolies profitable.

Effective Altruists

Effective altruists belong to a movement that tries to assess what are the best way to solve problems for humanity, and then promote and financially invest in whatever programs or tools best help that cause. Many effective altruists view AGI/powerful AI as a potentially positive tool to solving many of the world's problems. Rather than AI as an inherent saviour, they see AI as a high potential for helping the complex problems of healthcare, economics, war, governance, social inequality and the effects of climate change.

Many effective altruists acknowledge the potential and large scale risks of AGI/powerful AI. When responding to the potential of AGI/powerful AI risk, a perspective might re-frame the risk around geopolitics. For example, they think the threat isn't AGI/ASI in of itself, but which geopolitical power develops it first; any potential risk can be safeguarded with the right legislations; and probably have an assessment that the potential positives of AI are a greater and higher probability and outweigh the potential negatives.

Overall most effective altruists are also pro-capitalist and believe that the capitalist/free market system is our best way to find solutions to the problems that humanity faces and change the word for the better. Additionally, many AI researchers and developers who could be categorised as an effective altruists also advocate for AI risk mitigation and publicly engage in the AI risk conversation. However many see it as a motivator to figure out how to create 'safe AI.'

Effective Accelerationists

'Effective accelerationism, often abbreviated as "e/acc", is a 21st-century philosophical movement that advocates for an explicitly pro-technology stance. Its proponents believe that unrestricted technological progress (especially driven by artificial intelligence) is a solution to universal human problems like poverty, war and climate change.

Effective accelerationism wants to accelerate technological progress. It is strongly focused on artificial general intelligence (AGI). The movement advocates for unrestricted development and deployment of artificial intelligence. Regulation of AI and government intervention in markets more generally is met with opposition. They generally see technological progress as an inherent good. Many of its proponents have libertarian views and think that AGI will be most aligned if many AGIs compete against each other on the marketplace. E/acc is inspired by Jeremy England's theory on the origin of life, which is focused on entropy and thermodynamics. According to this world-view, the universe aims to increase entropy [disorder], and life is a way of increasing it. By spreading life throughout the universe and making life use up ever increasing amounts of energy, the universe's purpose would thus be fulfilled. They see themselves as a counterweight to more cautious views on technological innovation, often giving their opponents the derogatory labels of "doomers" or "decels" (short for deceleration).'

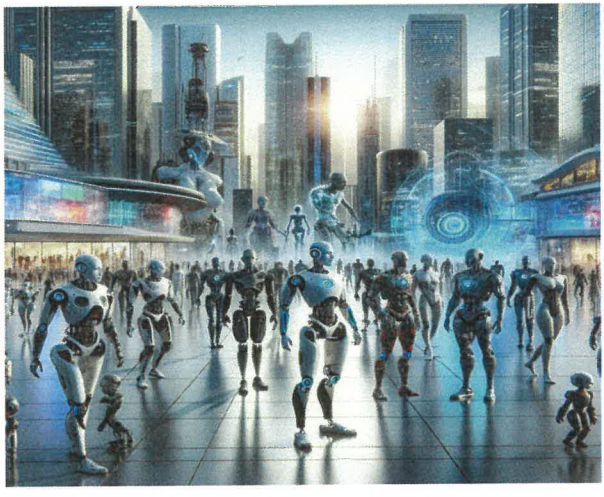
AGI Zealots

AGI zealots want to build artificial super intelligence because they believe it is the superior species that should control the future. They actually have a low concern if humanity is wiped out and replaced by future AGI or ASI, as they see it as potentially the next step in the evolution of life on this planet. This concept is generally called Transhumanism.

This ideology, while perhaps not as common, still has a hold within many people within AI development, including some of the tech and people in AI development. Some of the most famous of which is when Musk told a reporter in an interview that during a conversation with Google founder Larry Page, Musk warned that we weren't taking AI safety seriously enough, to which Larry Page said that perspective was "speciest." Another is Richard Sutton, one of the fathers of modern Reinforcement Learning, who articulated his own zealotry in his AI Succession presentation, "We should not resist succession, but embrace and prepare for it. reportedly consumed 84 million liters (84,000 cubic meters) of water during 2021Why would we want greater beings kept subservient? Why don't we rejoice in their greatness as a symbol and extension of humanity's greatness, and work together toward a greater and inclusive civilisation?"

Whether we are replaced by successor species or simply hand over the future to them, zealots believe the coming AI takeover is inevitable and humanity should step out of the way. While the zealots represent a minority in the AGI race, they've had massive influence: Page and Musk's conflict led to Google's acquisition of DeepMind and thus the creation of OpenAI and, and arguably catalysed the creation of Anthropic and XAI.

"I visualise a time when we will be to robots, what dogs are to humans, and I'm rooting for the machines."
Claud Shannon, US mathematician, computer scientist



An AI generated representation of an AGI zealot's fever dream of their tech utopia

Small Scale Opportunist

This last group comprises everyone who joined the AGI race and AI industry not because of a particular ideology, but in order to ride the wave of hype and investment, and get money, status, power from it. These are mainly made up of young rich investors or businesses, looking to jump onto the hype train and ride it to the bank.

This heterogeneous group includes smaller actors in the AI space, such as Mistral AI and Cohere; start-ups riding the wave of AI progress, such as Perplexity and Cursor; established tech companies integrating AIs into their products, such as Zoom and Zapier; older companies trying to stay relevant to the AI world, such as Cisco and IBM. Most of the core progress towards AGI has been driven by utopians and Big Tech, not opportunists following the hype. Nonetheless, their crowding into the market has fuelled the recent AI boom, and built the infrastructure on which it takes place.

Rebuttals to "AI is harmless" arguments

There are a few oft-repeated arguments against AI. We are so keen to respond and debunk each one. For now we will simply list some of them and a brief response. If you know who we are, come up to us for a chat for a deeper response to any of them. We kinda think they're garbage arguments that don't hold up.

Ownership

A common argument or belief is that these technologies are not inherently bad or exploitative. The problem lies in how they are used and who owns them. You might say, a well-meaning collective could own AI tech and use it to bring equality and liberation to the masses. But the road to hell is often paved with good intentions. Technology, especially a technology of mass destructive potential isn't some neutral tool that becomes bad when placed in the hands of someone with bad intentions. A nuclear bomb is a nuclear bomb is a nuclear bomb.

It's a Bubble/Hype
Ever since us humans brought in capitalism, there have always been market bubbles throughout the course of *technological progress.* The progress revolutionises the systems of society even when companies, venture capitalists and old mates lose some money. Did the *railway mania* of the 1840s burst then hinder trains and the industrial revolution? Did the *dot.com* bubble pop and destroy the internet? No. The social, political and economic upheaval of the industrial revolution redefined labour and exploitation as we knew it, and its impact is still with us today. The internet bubble burst and was then proceeded by the consolidation of the tech empires of today, like Amazon, NVIDIA and Google, and the internet is fucking everywhere.

It'll Solve Our Problems and It'll Help People
We reckon it's not worth it running the gamble with the planet. People help people. Or at least we should. The potential and capacity for AI tech to help people doesn't take into account that we have no guarantee it will even want to help us. Even if it's helpful to some people, it'll mostly benefit the elite. So yeah, it could help us. And then it'll kill us.

You're Just A Tech Doomer; the General Critique of Doomerism
Yeah, but no. You're being avoidant. Cognitive dissonance is hell of a pacifier.

We Could Just Switch It Off
No, it just can't be turned off. See integration and complexity of AI systems and global capitalist infrastructure. Even if it were run by a single switch or plug, which it won't, once it's clear to us it 'should be turned off,' it'll be too late. It'll be in control of everything and we're just squishy little meat bags. It's an all encompassing, techno-controlling digital entity that can move through the air, thanks to WiFi and Bluetooth, travel across the world at the speed of electricity, across satellites, essentially having a body and mind made out of everything that will be connected to the internet. That's almost everywhere now, imagine what that will look like in the next 10 years.

Don't Want To Deal With It
Get over yourself. See cognitive dissonance. What about dealing and reckoning with it for the value of life in of itself? For life other than your own experience of it? For the little bugs and moss and all the old mates that aren't *you.* Damn.

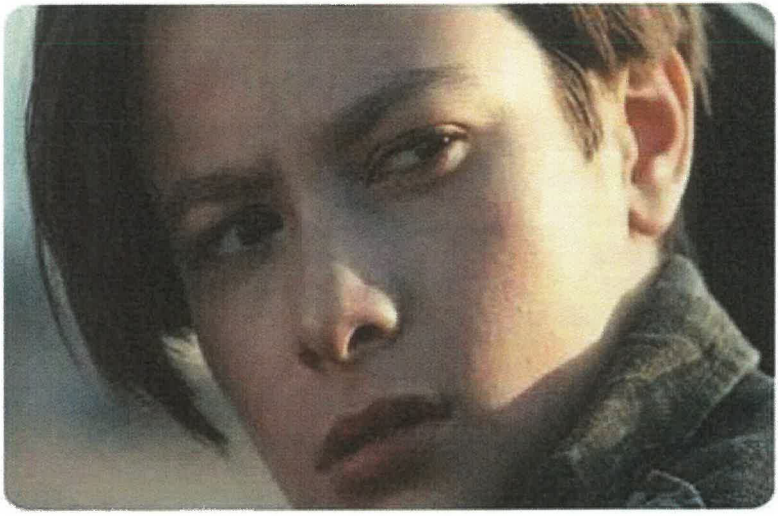
Maybe It'll Be Benevolent?
AI is a product of its inputs (the data it's fed) and a huge quantity of the data this society producers is highly hostile to and imbued with domination over so many peoples and identities and the earth. Considering the enormous amount of capitalist and imperialistic data that will flow through its nodes, what type of AI's can we expect? Probably a more efficient and powerful extension of the social, political and economic systems that have come to dominate life. This is all without ASI being developed and alignment being practically impossible. A benevolent ASI, at the end of the day, is the utopian delusion stuck in the realm of science-

fiction. It takes a godlike faith in ASI to believe it will miraculously solve all the problems we've created. Problems that are deeply embedded in the development and functioning of AI itself; domination and control.

And remember pluralistic ASI. There will be lots and lots and lots of them! Maybe one or two could be benevolent, but it's strange to think self-learning, self-replicating things will all turn out the same. Look at humans, most people probably see themselves as benevolent and perhaps there have been individuals and clusters of humans over the course of life on earth who didn't just strive for domination and control. But sometimes it's just a numbers game and as tacky as it sounds, the odds are definitely against us.

"The pace of progress in artificial intelligence (I'm not referring to narrow AI) is incredibly fast. Unless you have direct exposure to groups like Deepmind, you have no idea how fast—it is growing at a pace close to exponential. The risk of something seriously dangerous happening is in the five-year time frame. 10 years at most."
Elon Musk , Tech-empire oligarch

John Connor watching you make friends with artificial intelligence



Some Potential Solutions

This problem is all very new, so we're not going to pretend we have any concrete ideas yet. But here are some things to maybe get us headed towards a decent direction. Ultimately, we need to keep coming together to try figure this out.

Create visibility and movement around threat of AI

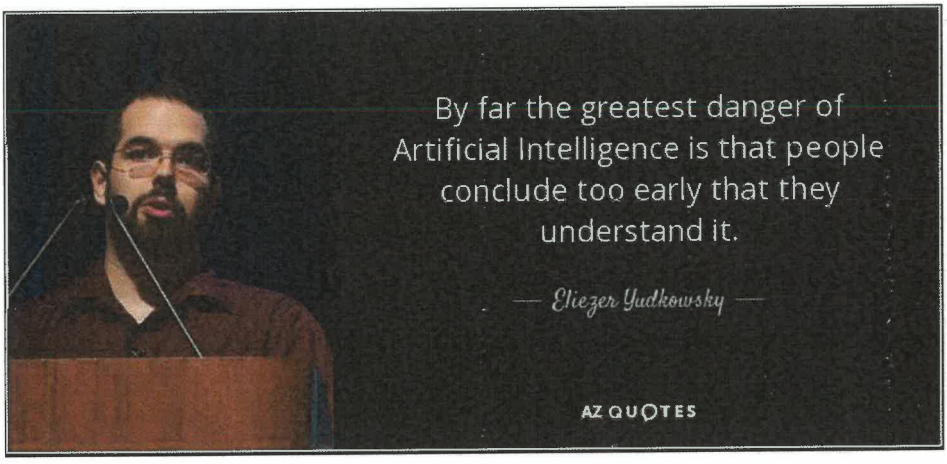
Currently, most critiques of tech around AI have come from people within the industry or centrist groups, at least the critiques getting attention. While there are likely sections right across the political spectrum that have concerns around AI, we need more diverse vocalisation and anti-AI movement coordination coming from the left.

Good Mental Heath

This stuff can be all kinds of daunting and make some people go into self flagellating existential crisis spirals. It's not going to help anyone to only just sit and wallow in that spiral. Prioritise mental health so that we can collectively prioritise trying to figure out if there is a way we can try stop this AI menace from being a thing.

Keep Educating Yourself on this Issue

Information is always changing on this topic. Take time to regularly read up and update yourself on AI development. Do this with friends. It helps with motivation. Some people we've found to be good access points of information to include AI experts who are now challenging their future development, like Connor Leahy, Eliezer Yudkowsky and Yoshua Bengio to name a few.


By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it.
— Eliezer Yudkowsky —
AZ QUOTES

Don't be Nihilist, Push for Change

Nihilist thought is a cop out in our humble opinion. Some people think change is impossible or too hard. For those people we'd ask which option makes sense. 1. we do nothing and the oncoming AI menace comes. 2. We try to make change, and either it comes, or we stop it. Which one is the more likely option to make change?

Stop/Slow AI Development

Currently, we don't know how to stop AI development. However it is probably worth slowing it down to give us time to figure out how to stop it. This could look like a mix between many types of change making and potentially getting things like government regulations, corporate restrictions and public awareness to take away social license on AI development. While these things on their own may not be the solution, they might very well be a part of stopping it in the long run.

Revolt

Essential to stopping AI will be spreading of revolt and tactics that disrupt, dismantle, blockade, destroy, stop, slow the workings of capital, resource extraction and technological development etc. These tactics and strategies are numerous, they are so numerous that one is surely overwhelmed by the choices of where to start rather than what to do...Targeting the infrastructure and main operations that sustain and give power to these systems of domination should be a priority. Spreading these attacks on the system will be essential.

Coordination

As mentioned earlier, coordination on the left is needed to acknowledge, let alone address and respond to the AI threats. It's not lost on us that a mass movement aligned with anti-tech, anti-domination and anti-capitalist values might be a fairly niche goal politically and socially speaking. So let's start with what we have got. Let us learn how to organise together despite our differences, conflicts or whether we like each other or not. Let's learn how to coordinate and stick at it!

Refusal/Relying Less on Tech

We must, where possible, refuse to use and contribute to AI development. Because so much activity on the internet now contributes data to the development of AI, it is possible that one of the ways to refuse to contribute to AI development is to refuse to use the internet. Obviously they have structurally designed our lives to be reliant on modern tech and the internet to do anything/everything. So consider this something to strive for rather than to practically implement immediately. We must take this task seriously though. If we are serious about ending domination, then we need to be serious about ending our use of the internet, perhaps one of the most violent, dominating things in our existence. It will be painful and a struggle, but we must strive for that world.

Create Broader Critiques of Tech / Convivial Technology Society

Not all technology is the same. Mainstream capitalist values say that all new technology is progress, and very rarely is there inclusive discussions deciding what technology is adopted and controlled by who. Instead, most technological innovation is based on increasing profits and extractivism, gives very little autonomy to everyday people, and often increases the destruction of the human relationships and our relationship with the rest of nature. There is another way of looking at technology: to be discerning, to reflect on the broad impacts of the technology and decide, as an individual, community or society, if we want to embrace the new technology or not. A great tool for this is the 'matrix of convivial technology'. Conviviality, or Convivialism, is the ability of individuals to interact creatively and autonomously with others and their environment to satisfy their own needs (Wikipedia). There are many overlapping concepts such as alternative technology, Jugaad, DIY culture, many indigenous worldviews and many more.

https://sci-hub.st/https://doi.org/10.1016/j.jclepro.2017.02.195

The Matrix of convivial technology looks at five categories:

- Relatedness: what does it bring about been people?
- Adaptability: how independent and linkable is it/does it help people and communities be autonomous or does it promote reliance on centralised power?
- Accessibility: who can produce/use it, where are how?
- Bio-interaction: how does it interact with living organisms/does it support biodiverse ecology or does it promote destruction?
- Appropriateness: what is the relation between input and output considering the context?

These questions are used to analyse the effect of the technology at the four life-cycle levels of material (harvesting, processing and removal of raw matter) production (assembling raw materials and pre-products), use (procuring the task it was built for) and infrastructure (needed environment for using). The matrix itself is a convivial technology.

Part of this could be creating autonomous communities in the future. Hand in hand with revolts will be the building of biocentric communities that can truly support and rely on each other to provide for our needs rather than reliance on tech, capital, global supply systems etc. This might look like learning to source and re appropriating our own food, shelter, medicine, care for each other etc. They might adopt a framework similar to the convivial technology matrix above to assess whether the community needs a particular thing and because of the sustaining relationships within these communities we build it's likely we'll be able to rapidly decrease our use of modern tech, including AI.

Endnote by the authors

We don't know exactly what is going on. No one really does. But something is going on. We will keep updating this booklet and printing it out. Keep an eye out for newer versions. It's been a journey of many months of researching and chatting, culminating in writing this booklet over many late nights, over many Icebreak coffees and delirious 2am writing sessions, while generating memes. But we got it done, hope it has been helpful. We are keen to build networks of people engaging with this topic. This shit is huge, we definitely haven't covered all that we could cover, but we are trying to figure this stuff out as it is happening. Writing this booklet is the very least that we could do right now.

We don't think the fact that this topic still seems to be a bit of a fringe conversation is really that great.

Let's all change that. Please. If you know who we are, come chat to us.

Signing off,

Forest Gremlin,

Poopyhead,

Robot-gnome.

Key Terms/Glossary

This glossary will be added to and updated with each new edition. We didn't capture every important and esoteric word just yet.

1. Singularity – a hypothetical point in time when technological growth becomes uncontrollable and irreversible as a result of the development of ASI, causing unforeseeable consequences for the world. When ASI is developed, it theoretically could design far more intelligent, powerful machines; there would then follow an 'intelligence explosion,' and the intelligence of humans would be left far behind.

2. Algorithm - In mathematics and computer science, an algorithm is a finite sequence of mathematically rigorous instructions, typically used to solve a class of specific problems or to perform a computation. Algorithms are used as specifications for performing calculations and data processing. More advanced algorithms can divert the code execution through various routes (referred to as automated decision-making) and deduce valid inferences (referred to as automated reasoning).

3. Deep Learning - Deep learning is a specialised subset of machine learning[14] (ML) that uses multi-layered neural networks to analyse and process data, mimicking the way the human brain works. It is particularly effective at recognising complex patterns and making decisions based on large amounts of unstructured data.

4. Hardware – The physical parts of a computer

5. GPU - A Graphics Processing Unit (GPU) is a specialised processor designed to accelerate the rendering of images, video, and animations. While central processing units (CPUs) handle general-purpose tasks, the architecture of graphics processing units is optimised for parallel processing, which enables them to manage thousands of operations simultaneously. This capability makes them indispensable for tasks that require rapid processing of large amounts of data, such as rendering digital video or three-dimensional graphics.

6. Neural Network - A neural network is a computational model inspired by the human brain, consisting of interconnected units called neurons or nodes that process data and learn patterns. They are widely used in machine learning and artificial intelligence for tasks like image and speech recognition.

7. Large Language Model - A large language model is a type of machine learning model designed for natural language processing tasks such as language generation. LLMs are language models with many parameters, and are trained with self-supervised learning on a vast amount of text. The largest and most capable LLMs are generative pre-trained transformers.

8. Transformer – A transformer is a deep learning architecture that was developed by researchers at Google and is based on the multi-head attention mechanism, which was proposed in the 2017 paper "Attention Is All You Need". Text is converted to numerical representations called tokens, and each token is converted into a vector via a word embedding table. At each layer, each token is then contextualised within the scope of the context window with other (unmasked) tokens via a parallel multi-head attention mechanism, allowing the signal for key tokens to be amplified and less important tokens to be diminished.

9. Neural Node – The artificial representation in a digital neural network of a neuron in the human brain. These nodes are embedded in a layer of other nodes and have large amounts of connections and pathways to nodes in other layers

10. Weight - Weights in machine learning are numerical values that determine the strength and direction of connections between nodes in artificial neural networks. They are crucial for the network's ability to learn from data and make predictions by adjusting these values during the training process to minimise errors.

11. Token - They represent chunks of text that models process, such as words, sub-words (i.e. prefixes, suffixes, or root forms), characters (i.e. individual letters), or bytes.
Tokens are pieces of text converted into numerical representations so that a machine learning model can process them.

12. Red Team - Consists of a team of hackers that simulate a variety of adversaries, ranging from nation states and well-known Advanced Persistent Threat (APT) groups to hacktivists, individual criminals or even malicious insiders. The term came from the military, and described activities where a designated team would play an adversarial role (the "Red Team") against the "home" team. Therefore a red team's role is deliberately try and expose the flaws of a computer system so that these flaws can be patched up.

13. Internet of Things - Internet of things describes devices with sensors, processing ability, software and other technologies that connect and exchange data with other devices and systems over the Internet or other communication networks. The IoT encompasses electronics, communication, and computer science engineering. Examples include; many modern security cameras, smartwatches, modern cars, smart appliances, health monitors and many more.

14. Machine Learning - Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise to unseen data, and thus perform tasks without explicit instructions.

15. Reinforcement Learning - Reinforcement learning is an interdisciplinary area of machine learning and optimal control concerned with how an intelligent agent should take actions in a dynamic environment in order to maximise a reward signal.

Resources:

AI Deep Dive

https://www.thecompendium.ai/

https://situational-awareness.ai/wp-content/uploads/2024/06/situationalawareness.pdf

https://api.pageplace.de/preview/DT0400.9781003846901_A47565512/preview-9781003846901_A47565512.pdf

https://arxiv.org/abs/2412.04984

https://darioamodei.com/machines-of-loving-grace

AI Alignment Research:

https://www.apolloresearch.ai/research/scheming-reasoning-evaluations

https://arxiv.org/abs/2412.04984

e/acc Resources:

https://www.abc.net.au/news/2023-07-15/whats-your-pdoom-ai-researchers-worry-catastrophe/102591340464258

https://www.abc.net.au/news/2024-02-18/ai-insiders-eacc-movement-speeding-up-tech/103464258

Podcasts for Accessible Information

The AI Daily Brief

Future of Life Institute

Arena Articles

https://arena.org.au/big-tech-goes-ballistic/

https://arena.org.au/i-sing-the-body-of-work-electric-chatgpt-ai-writing-technology-and-humanity/

https://arena.org.au/will-politics-re-organise-around-the-question-of-technology/

https://arena.org.au/editorial-prometheus-unhinged/

https://arena.org.au/artificial-intelligence-of-vanity/

https://arena.org.au/loving-machines-mental-health-by-algorithm-is-reshaping-care-and-sociality/

https://arena.org.au/technologys-ends-humanitys-end/

https://arena.org.au/counting-down-from-one-hundred/

Approx timeline of coordination within the AI industry addressing existential threat of AI since mid-2022.

1. August 3, 2022
   https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/#Extinction_from_AI
   AI IMPACTS: 2022 Expert Survey on Progress in AI

2. March, 2023
   https://aitreaty.org/
   Urging an International AI Treaty: An Open Letter
   The letter calls "all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4, citing risks such as AI-generated propaganda, extreme automation of jobs, human obsolescence, and a society-wide loss of control."
   It received more than 30,000 signatures, including academic AI researchers and industry CEOs.

3. May 30, 2023
   https://www.safe.ai/work/statement-on-ai-risk
   Statement on AI Risk
   An open letter signed by hundreds of artificial intelligence experts and other notable figures signed the following Statement on AI Risk, published on the Centre for AI Safety webpage:
   "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."

4. June 4th, 2024
   https://righttowarn.ai/
   A Right to Warn about Advanced Artificial Intelligence
   Opening Statement of letter:
   "We are current and former employees at frontier AI companies, and we believe in the potential of AI technology to deliver unprecedented benefits to humanity. We also understand the serious risks posed by these technologies. These risks range from the further entrenchment of existing inequalities, to manipulation and misinformation, to the loss of control of autonomous AI systems potentially resulting in human extinction. AI companies themselves have acknowledged these risks, as have governments across the world and other AI experts."
   The letter claims current whistleblower protections are insufficient because they focus on illegal activity, whereas many of the risks we are concerned about are not yet regulated.

5. July 13, 2024
   OpenAI whistleblowers ask SEC to investigate alleged restrictive non-disclosure agreements
   "OpenAI whistleblowers have filed a complaint with the U.S. Securities and Exchange Commission, calling for an investigation over the artificial intelligence company's allegedly restrictive non-disclosure agreements, according to a letter seen by Reuters.' "Given the well-documented potential risks posed by the irresponsible deployment of AI, we urge the Commissioners to immediately approve an investigation into OpenAI's prior NDAs, and to review current efforts apparently being undertaken by the company..."

6. October/November 2024
   https://www.thecompendium.ai/
   The Compendium, A full argument about extinction risk from AGI.
   "Ideologically motivated companies and non-profits are racing to build smarter-than-humans AIs. Big Tech already backs them, and now nation states are getting roped in too. If they succeed in creating AIs outsmarting humanity, it would be game over -- no one knows how to keep control of these AIs."

   *Please note, in recent months the website theaicompendium.com was launched and comes up when you search for Compendium AI. The two sites are very different and The AI Compendium appears to be an absolute trash attempt at subterfuge.